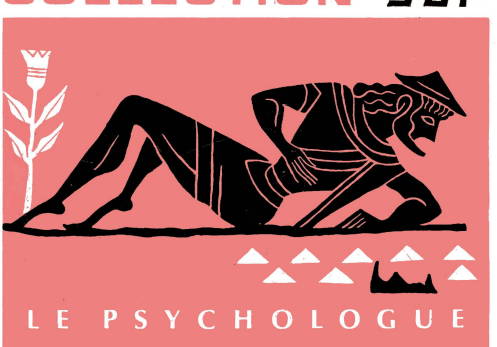
MARC BARBUT

MATHÉMATIQUES *DES* SCIENCES HUMAINES

II. *Nombres et Mesures*

COLLECTION SUP



PRESSES UNIVERSITAIRES DE FRANCE

DU MÊME AUTEUR

Mathématiques des sciences humaines, I — *Combinatoire et Algèbre*, Paris, Presses Universitaires de France, 1967.

« LE PSYCHOLOGUE »
Section dirigée par Paul FRAISSE
Professeur à la Sorbonne

33

MATHÉMATIQUES DES SCIENCES HUMAINES

II — *Nombres et Mesures*

par

MARC BARBUT
Directeur d'Études
à l'École pratique des Hautes Études (6e Section)
Professeur associé à la Faculté des Lettres et Sciences humaines
de Paris-Sorbonne

PRESSES UNIVERSITAIRES DE FRANCE
108, Boulevard Saint-Germain, Paris
1968

DÉPOT LÉGAL

1re édition .. .. .. 1er trimestre 1968

TOUS DROITS

de traduction, de reproduction et d'adaptation réservés pour tous pays

© 1968, *Presses Universitaires de France*

AVANT-PROPOS

Ce manuel complète le cours de mathématiques de première année du premier cycle en psychologie et sociologie (chap. I à VI inclus), dont la majeure partie a été publiée dans *Mathématiques des Sciences humaines*, I : *Combinatoire et Algèbre*, Presses Universitaires de France, 1967 (ouvrage qui est noté *MSH* dans le présent manuel) ; en outre, il aborde très largement le cours de calcul des probabilités de seconde année (chap. VII à IX) ; bien que très élémentaire, il est certainement d'une lecture plus ardue que *MSH* pour les étudiants auxquels il s'adresse.

L'idée qui a présidé à sa rédaction a été de regrouper dans un seul volume l'essentiel des outils mathématiques qui servent dans la plupart des techniques de *mesure* utilisées par les Sciences humaines, mais qui sont *préalables* à ces méthodes ; en particulier, on s'est efforcé d'y fournir les bases mathématiques qui permettront de comprendre ultérieurement, et d'utiliser correctement les techniques d'inférence statistique. Bien entendu, comme il s'agit ici d'un cours de mathématiques et non de méthodologie, ni les méthodes (statistiques ou non) de mesure elles-mêmes, ni les applications ne sont abordées. Mathématiques des Sciences humaines ne veut pas dire Mathématiques *appliquées* aux Sciences humaines, mais Mathématiques *qui s'appliquent* dans ces Sciences.

On a donc commencé par introduire très (trop) succinctement les structures numériques fondamentales (échelles quantitatives, ou numériques) et les fonctions qui leur sont associées (première partie) ; puis, on a mis l'accent sur la machinerie dont relèvent la plupart des techniques de mesure par des nombres, en description statistique comme en calcul des probabilités : *l'algèbre linéaire* (dont l'introduction a été faite dans *MSH*), et plus précisément les *formes linéaires et bilinéaires* (produit scalaire) sur un vectoriel. C'est ainsi, par exemple, que les axiomes classiques du

calcul des probabilités (axiomes de Kolmorogoff) équivalent à la donnée d'une forme linéaire positive sur un vectoriel d'aléas numériques (chap. IX, p. 225).

Certaines techniques non linéaires, ainsi que les échelles ordinales (types d'ordre), ont néanmoins été abordées ; la place a manqué pour traiter de la structure de *treillis distributif* qui joue, par rapport à beaucoup de ces techniques, le rôle que joue celle de vectoriel par rapport aux techniques numériques.

On est constamment resté dans le cas *fini* (vectoriels de dimension finie ; probabilités sur des ensembles finis), de façon à ne pas voiler les idées par des difficultés spécifiques à l'introduction de l'infini (passages à la limite et questions d'existence) ; l'étudiant sera ainsi mieux armé, semble-t-il, pour aborder plus tard les mesures de probabilité sur des ensembles infinis souvent utilisées, à titre d'approximation commode, en inférence statistique.

Quant à l'inférence statistique elle-même, on s'est contenté dans ce manuel d'essayer d'en faire pressentir la logique, à la fin de l'ouvrage : un modèle relevant du « calcul des probabilités », appliqué à une description statistique correcte des observations : la première loi des grands nombres (de Bernoulli) fournit l'occasion de ce contact préliminaire avec la statistique inductive, plus précisément avec le problème de l'estimation.

Ce manuel doit beaucoup aux critiques et suggestions que m'ont faites des collègues participant à l'enseignement des mathématiques et de la statistique dans les Facultés des Lettres et Sciences humaines, et tout particulièrement à celles d'Henri Rouanet pour la deuxième partie.

D'autre part, J.-L. Piednoir, assistant de Mathématiques à la Faculté des Lettres et Sciences humaines de Paris, a participé à l'élaboration du manuel, en rédigeant notamment le § 8 du chapitre I^er^, et une grande partie du chapitre II.

M. B.

SOMMAIRE

PAGES

AVANT-PROPOS 1

INSTRUCTIONS DE LECTURE POUR LES ÉTUDIANTS 5

BIBLIOGRAPHIE 7

PREMIÈRE PARTIE

LES NOMBRES

CHAPITRE PREMIER. — *Nombres, ordres, approximations*.... 11

— II. — *Fonctions numériques* (avec la collaboration de J.-L. PIEDNOIR) 38

DEUXIÈME PARTIE

LA MESURE EN DESCRIPTION STATISTIQUE

CHAPITRE III. — *Produit scalaire* 81

— IV. — *Mesures de la distance et de la dispersion*. 110

— V. — *Cosinus et corrélation linéaire* 138

TROISIÈME PARTIE

LA MESURE EN PROBABILITÉS

CHAPITRE VI. — *Mesures de probabilité* 163

— VII. — *Aléas* 192

— VIII. — *Probabilités conditionnelles* 213

— IX. — *Aléas numériques* 241

INDEX DES SYMBOLES ET DES NOTATIONS 283

INDEX DES TERMES 285

INSTRUCTIONS DE LECTURE POUR LES ÉTUDIANTS

1) Les *paragraphes marqués d'un astérisque* * peuvent être omis en première lecture.

2) Les *renvois en cours de texte* concernent :
   — un paragraphe du même chapitre s'ils comportent seulement la mention d'un paragraphe : § *x* ;
   — un paragraphe d'un autre chapitre, s'ils comportent la mention d'un chapitre et d'un paragraphe : chap. *x*, § *y* ;
   — un paragraphe de *Mathématiques des Sciences humaines*, I — *Combinatoire et Algèbre* (P.U.F., 1967), s'ils sont notés : *MSH*, chap. *x*, § *y*.

3) *Deux index*, à la fin du manuel, fournissent la liste (alphabétique) des symboles et notations et celle des termes définis dans le texte ; il faut s'en servir.

4) $f(x)$ *et* fx sont les deux façons qui ont été adoptées dans le manuel pour noter l'image, dans une application f, d'un élément x de l'ensemble de départ de celle-ci ; en particulier, on a utilisé la notation $f(x)$ au chap. II, et fx (pour éviter des surcharges de parenthèses) au chap. IX.

5) Les notions fondamentales d'application (fonction), d'ensemble, parties d'ensemble et leur algèbre, vecteurs et leur algèbre, constamment utilisées dans ce manuel, sont supposées connues.

6) Refaites vous-même figures, calculs et démonstrations ; n'omettez pas les exercices que comporte le manuel.

BIBLIOGRAPHIE

1. Recueils d'exercices corrigés

Cahiers mathématiques, 1 : *Exercices corrigés sur des structures élémentaires* (Mouton & Gauthier-Villars, 1966). Exercices sur les trois parties du manuel.

Cahiers mathématiques, 2 : à paraître (1968) (Gauthier-Villars & Mouton). Exercices sur la première partie.

E. Berrebi, *Exercices corrigés de mathématiques* (Dunod), t. 1 et 2. Première et deuxième parties.

C. Labrousse, *Statistique*, t. 1 (Dunod, 1965). Exercices avec corrigés pour les étudiants en Sciences économiques.

2. Tables numériques

Tables statistiques du Centre de Formation aux Applications Industrielles de la Statistique, I.S.U.P., 9, quai Saint-Bernard, Paris (V^e^). Tables des principales fonctions utilisées en statistique ; en particulier, table de carrés, tables de distributions binomiales. (Troisième partie du manuel.)

Tables numériques de fonctions élémentaires, par J. Laborde (Dunod) : fonctions, puissances, exponentielles, logarithmes, trigonométriques ; nombres binomiaux. (Chap. II, VII et IX du manuel.)

3. Compléments au manuel

Première partie

G. Th. Guilbaud, *Mathématiques* (Presses Universitaires de France). Seconde partie : Fonctions exponentielles et logarithmes.

R.-V. Andree, *Initiation à l'algèbre* (Gauthier-Villars & Mouton, 1968) : groupes, anneaux, corps, structures numériques.

Deuxième partie

L. Guerber et P.-L. Hennequin, *Initiation à la statistique* (Bibliothèque d'Enseignement Mathématique, A.P.M.E.P., 29, rue d'Ulm, Paris, V^e^).

G. Calot, *Cours de statistique descriptive* (Dunod).

Ces deux ouvrages contiennent de nombreux exercices.

Deuxième et troisième partie

M. GIRAULT, *Éléments de méthodologie statistique* (Dunod).

R. DAVAL et G. Th. GUILBAUD, *La méthode statistique*, 2 vol. (Presses Universitaires de France).

Troisième partie

P. ROSENSTIEHL, *Mathématiques de l'action* (Dunod).

D. HERAULT, *Éléments de théorie moderne des probabilités* (Dunod).

4. NOMBRES ET MESURES EN MÉTHODOLOGIE DES SCIENCES HUMAINES

Les structures numériques ou ordonnées sont partout présentes en méthodologie des Sciences humaines ; on trouvera donc dans tous les ouvrages sur les méthodes des applications des questions traitées ici. Comme ouvrages d'ensemble, on peut recommander :

— en psychologie expérimentale :

M. REUCHLIN, *Les méthodes quantitatives en psychologie* (Presses Universitaires de France).

— en sociologie :

R. BOUDON, *Analyse mathématique des faits sociaux* (Plon).

PREMIÈRE PARTIE

LES NOMBRES

CHAPITRE PREMIER

NOMBRES — ORDRES APPROXIMATIONS

1. ENSEMBLE DE NOMBRES

a) *On mesure avec des nombres*

La *mesure* de « grandeurs » ou de « variables » qui interviennent dans les sciences physiques ou humaines se fait au moyen des nombres.

Mesurer, c'est d'abord affecter à chacun des objets que l'on mesure un nombre et un seul, sa mesure ; c'est définir une *application* de l'ensemble des objets que l'on mesure dans un ensemble de nombres.

Les nombres qui mesurent les objets servent à *comparer*, à *calculer* ; en effet, deux nombres sont toujours comparables entre eux, et on peut les additionner, parfois les soustraire l'un de l'autre, les multiplier entre eux, parfois les diviser l'un par l'autre. Dans les sciences humaines, on parle souvent, à leur propos, d'*échelles quantitatives* ; dans cette expression, *échelle* signifie qu'il y a un *ordre* entre nombres (la comparaison) et *quantitative* signifie que l'on peut calculer, en particulier par addition.

Par exemple, les longueurs se comparent et s'additionnent entre elles ; il en est de même des temps de réaction de sujets à des stimulus en psychologie expérimentale.

Dans la seconde et la troisième partie de ce cours, on

étudiera le statut précis de la mesure en mathématiques ; préalablement, puisqu'on mesure avec des nombres, il est nécessaire d'examiner ce que sont les nombres eux-mêmes.

b) *Les principales espèces de nombres*

Et d'abord, il y a plusieurs espèces de nombres, plusieurs ensembles de nombres ; voici la nomenclature de ceux qui servent pour la mesure.

Les entiers naturels : 0, 1, 2, ..., n, ..., dont l'ensemble est noté **N** :

$$\mathbf{N} = \{ 0, 1, 2, 3, \ldots, n, \ldots \}$$

Les entiers relatifs, c'est-à-dire positifs ou négatifs, ..., — 2, — 1, 0, 1, 2, ..., dont l'ensemble est noté **Z** :

$$\mathbf{Z} = \{ \ldots, -2, -1, 0, 1, 2, \ldots \}$$

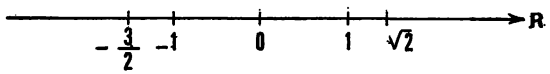
Les rationnels positifs (fractions *positives*, 0 exclu), tels que 1/2, 3/5, etc., dont l'ensemble est noté $\mathbf{Q}^*$ (lire **Q** étoile). $\mathbf{Q}^*$ est l'ensemble des fractions p/q, où p et q sont des *entiers naturels* différents de 0. Dans cette définition, les fractions sont supposées réduites à leur plus simple expression ; par exemple, 2/3 et $6/9 = \frac{2 \times 3}{3 \times 3}$ représentent le même nombre rationnel 2/3.

Les rationnels (fractions positives, négatives ou nulles), tels que — 1/3, — 4/5, 0, 6/7, etc., dont l'ensemble se note **Q**. C'est l'ensemble des fractions p/q, où p et q sont des entiers *relatifs*, et où $q \neq 0$. Parmi les rationnels il y a, en particulier, les nombres entiers relatifs : fractions dont le dénominateur vaut 1 ; par exemple : $-4/1 = -4$; $7/1 = 7$.

Les réels positifs, dont l'ensemble est noté $\mathbf{R}^*$, qui comprennent, outre les rationnels *positifs*, tous les nombres dits *irrationnels* positifs, comme $\sqrt{2}$, π *(pi)*, $\sqrt[3]{7}$, etc.

Les réels, dont l'ensemble est noté **R**, qui comprennent, outre les rationnels, tous les *irrationnels* positifs ou négatifs : $-\sqrt{2}$, $-\sqrt[3]{9}$, $\sqrt[5]{7}$, etc.

On se représente souvent l'ensemble des réels comme l'ensemble des points d'une droite orientée, sur laquelle ont été choisies une origine et une unité :



On note les inclusions :

$$\mathbf{R}^* \supset \mathbf{Q}^*$$
$$\mathbf{R} \supset \mathbf{Q}$$

Les *décimaux*, dont l'ensemble est noté **D** ; ce sont tous les nombres positifs ou négatifs, qui, en *écriture décimale*, s'écrivent avec un nombre fini de chiffres : 3,14 est un nombre décimal ; mais le rationnel $1/3 = 0{,}333...$, qui comporte une infinité de fois le chiffre 3, n'est pas un décimal.

Tout décimal est un rationnel, car égal à une fraction :

$$3{,}14 = \frac{314}{100}$$

Par contre, comme on vient de le voir sur l'exemple du rationnel 1/3, tout rationnel n'est pas un décimal. On a l'inclusion :

$$\mathbf{D} \subset \mathbf{Q}$$

Comme chaque nombre entier est un décimal, on a finalement entre les ensembles **N**, **Z**, **D**, **Q**, **R**, la suite d'inclusions :

$$\mathbf{N} \subset \mathbf{Z} \subset \mathbf{D} \subset \mathbf{Q} \subset \mathbf{R}$$

Il y a encore d'autres sortes de nombres ; dans ce cours, seuls les ensembles dont on vient de donner la liste serviront.

Pour chacune de ces espèces, les règles de calcul et la structure d'ordre ont des propriétés spécifiques ; il *ne faut jamais parler de nombres sans préciser desquels il s'agit.*

2. Entiers naturels : N

a) *Un ordre total discret*

Les entiers naturels :

$$\mathbf{N} = \{ 0, 1, 2, \ldots, n, \ldots \}$$

constituent d'abord un ensemble infini (1) ; la suite des entiers naturels est illimitée ; après chaque entier naturel, il y a son suivant : après 0 vient 1, après 1 vient 2, ..., après 3 514 vient 3 515, etc.

La *structure d'ordre* de N est définie par cet ordre de consécution ; on note $x < y$ si l'entier x vient avant l'entier y ; on dit que x est avant y, ou encore que x est *plus petit* que y.

On a ainsi :

$$0 < 1 < 2 < \ldots < n < (n + 1) < \ldots$$

La relation $<$ entre entiers naturels est bien une *relation d'ordre* dans le sens donné à ce mot en mathématiques (cf. *MSH*, chap. IX) : elle est *transitive*, et *antisymétrique* ; de plus, elle est *totale* : deux entiers sont toujours comparables.

En outre, cet ordre est discret, ce qui signifie qu'entre deux éléments il y a toujours un nombre *fini* d'*intermédiaires* (contrairement à ce qui se passe par exemple pour l'ordre des points sur une droite : entre deux points, il y a une *infinité* d'intermédiaires). En particulier, entre deux entiers consécutifs, tels que 2 et 3, il n'y a *aucun intermédiaire.*

Techniquement, on dit, selon la terminologie de G. Cantor, que l'ordre de N, qui est total, pour lequel il y a un plus petit élément, 0, et pour lequel entre deux éléments

(1) N désigne parfois aussi l'ensemble des entiers strictement positifs, 0 exclu : 1, 2, ...

il y a toujours un nombre fini d'intermédiaires, est de type ω *(oméga)* (cf. § 7).

On note (N, $>$) la *structure d'ordre* de N.

b) *Une arithmétique*

D'autre part, les entiers naturels peuvent toujours être *additionnés* et *multipliés* entre eux :

$$x + y \qquad x \times y$$

Ces opérations définissent le *calcul arithmétique*, dont les règles sont bien connues, et dont certaines valent pour les deux opérations :

$+$		$\times$
$x + y = y + x$	commutativité	$x \times y = y \times x$
$x + (y + z) = (x + y) + z$	associativité	$x \times (y \times z) = (x \times y) \times z$
$0 : x + 0 = x$	élément neutre	$1 : x \times 1 = x$

D'autres règles ne valent que pour l'une des opérations seulement : celles qui mettent en jeu le passage à l'opération inverse, soustraction pour $+$, division pour $\times$.

On ne peut *soustraire* l'entier naturel y de l'entier naturel x, que si $x \geqslant y$ (1) et le résultat est un entier naturel : $5 - 3 = 2$.

Mais $(3 - 5)$ n'est pas défini dans N (ce serait un entier négatif).

De même, on ne peut toujours *diviser* x par y de façon à obtenir un quotient x/y qui soit un entier naturel ; mais la condition pour que la division soit possible n'est plus $x \geqslant y$, mais que x soit un multiple de y, et y distinct de 0 : 28 est divisible par 4, et le quotient est 7 ($28 = 4 \times 7$) ; mais 28 n'est pas divisible par 5.

(1) $x \geqslant y$: x supérieur *ou égal* à y.

On note (**N**, +) la structure de **N** par rapport à l'*addition* ; comme cette opération est associative, et a un élément neutre o, on dit que (**N**, +) est un *monoïde*. Comme l'opération + est en outre commutative, on dit que c'est un monoïde *commutatif* (cf. *MSH*, chap. II, § 5).

De même, la structure de **N** par rapport à la multiplication, notée (**N**, ×) est une structure de monoïde commutatif.

Le calcul par addition et le calcul par multiplication sont reliés entre eux par la propriété de distributivité de la multiplication par rapport à l'addition.

$$(x + y) \times z = x \times z + y \times z$$

c) *L'ordre et l'arithmétique sont compatibles entre eux*

Le calcul par addition et le calcul par multiplication sont également reliés à la *structure d'ordre* : l'addition est *compatible* avec l'ordre, en ce sens qu'ajouter un même entier naturel à deux autres ne change pas l'ordre de ceux-ci :

$$x > y \Rightarrow x + z > y + z$$

Il en est de même pour la multiplication, pourvu que l'on multiplie par un entier naturel différent de zéro :

$$x > y \Rightarrow [z \neq 0 \Rightarrow x \times z > y \times z]$$

En résumé, deux calculs, par + et par ×, avec leurs analogies et leurs différences ; un lien entre ces deux calculs (distributivité) ; un lien entre chacun d'entre eux et l'ordre des entiers naturels.

3. Entiers positifs et négatifs : Z

a) *Un ordre total discret*

Comme **N**, l'ensemble **Z** est le support de plusieurs structures : d'abord c'est un ensemble *ordonné* par la relation >, qui se lit « être plus grand que » :

$$6 > -2 > -3 ; \quad 3 > 1 > -2 ; \quad \text{etc.}$$

Cet ordre est *total* : deux entiers sont *toujours comparables*.

Entre deux éléments quelconques il n'y a qu'un nombre fini d'intermédiaires ; mais il n'y a pas de plus petit élément, comme dans $(\mathbf{N}, >)$; chaque élément a un précédent et un suivant : le précédent de 3 est 2, son suivant 4.

On note $(\mathbf{Z}, >)$ la *structure d'ordre* de **Z**. On dit que $(\mathbf{Z}, >)$ a le type d'ordre ζ *(dzeta)* (cf. § 7).

$(\mathbf{Z}, >)$ est un ordre total, tel qu'entre deux éléments il y ait toujours un nombre fini d'intermédiaires ; c'est un ordre *discret*.

b) *Un groupe abélien*

Par rapport à l'opération d'addition, $+$, **Z** constitue *un groupe* noté $(\mathbf{Z}, +)$; en effet :

— $+$ est *associative* : $x + (y + z) = (x + y) + z$;
— il y a un *élément neutre*, 0 : $x + 0 = 0 + x = x$;
— chaque élément x a un *opposé* $(-x)$:

$$x + (-x) = 0$$

De plus l'addition est *commutative* : $x + y = y + x$.

On dit que $(\mathbf{Z}, +)$ est un *groupe commutatif* (syn. : groupe *abélien*).

Tous les groupes ne sont pas commutatifs ; par exemple, le groupe des permutations d'un ensemble ne l'est pas (cf. *MSH*, chap. II).

$(\mathbf{Z}, +)$ est un groupe abélien (commutatif).

Comme $(\mathbf{Z}, +)$ est un groupe, on peut toujours y calculer des différences : $x - y$ est toujours défini ; c'est un entier positif ou négatif.

c) *Un groupe abélien ordonné*

La structure d'ordre de **Z** est *compatible* avec celle de groupe abélien, en ce sens qu'additionner un même entier à deux autres ne change pas l'ordre de ces deux entiers :

Si $x > y$, alors $\forall z$, $x + z > y + z$.

On résume cette propriété en disant que :

> $(\mathbf{Z}, +, >)$ est un groupe abélien ordonné.

d) *Un anneau*

Par rapport à la *multiplication* $\times$, **Z** n'est pas un groupe : $\times$ est associative, elle a un élément neutre 1, mais en général il n'y a pas d'élément « opposé » à un élément x, c'est-à-dire un entier x' tel que :

$$x \times x' = 1$$

Exercice :

Trouver tous les couples d'entiers (x, x') tels que $x \times x' = 1$. (Rép. : $x = 1$, $x' = 1$; $x = -1$, $x' = -1$).

$(\mathbf{Z}, \times)$ est une structure associative, munie d'un élément neutre : c'est un *monoïde.* Ce monoïde est d'ailleurs *commutatif :*

$$x \times y = y \times x$$

$(\mathbf{Z}, \times)$ diffère également de $(\mathbf{Z}, +)$ dans son comportement par rapport à l'ordre ; ici, il n'y a en effet plus compatibilité entre l'opération (la multiplication) et l'ordre ; multiplier deux mêmes entiers par un troisième peut changer leur ordre :

$2 < 4$ mais $2 \times (-3) = -6 > 4 \times (-3) = -12$

Par contre ordre et multiplication sont compatibles si l'on se restreint aux seuls entiers *positifs*.

Z, muni de son addition et de sa multiplication, noté $(\mathbf{Z}, +, \times)$, constitue un *anneau*.

Anneau : Tout ensemble dans lequel sont définies *deux* opérations, qui est un *groupe abélien* pour la première, un *monoïde* pour la seconde, et tel que la seconde opération soit distributive par rapport à la première opération.

Ici nous avons bien distributivité de $\times$ par rapport à $+$:

$$x \times (y + z) = x \times y + x \times z$$

$(\mathbf{Z}, +, \times)$ est un *anneau*

Avec en plus son ordre tel que :

$$x > y \Rightarrow x + z > y + z$$
$$x > y, z > 0 \Rightarrow x \times z > y \times z$$

c'est un *anneau ordonné*.

4. Les rationnels : $\mathbf{Q}^*$ et $\mathbf{Q}$

a) *Groupes abéliens*

De même que l'on a introduit les entiers négatifs pour pouvoir toujours calculer les différences $x - y$, de même on introduit les *rationnels* positifs et négatifs, c'est-à-dire les fractions, de façon à disposer d'une structure dans laquelle on puisse toujours calculer les quotients x/y : on a ainsi l'ensemble **Q** des rationnels :

$$\mathbf{Q} = \{ p/q,\ p \text{ et } q \text{ entiers},\ q \neq 0 \}$$

L'ensemble **Q** est à son tour support de structures : $(\mathbf{Q}, +)$ est un groupe abélien.

Muni de la multiplication, **Q**, noté $(\mathbf{Q}, \times)$, est un groupe abélien, pourvu que l'on en *supprime zéro* ; en effet, toute

fraction, sauf zéro, possède une fraction inverse ; l'inverse de p/q est q/p : $(p/q) \times (q/p) = 1$.

L'inverse de 0 n'existe pas puisque : $\forall a,\ a \times 0 = 0$; d'où, $\forall a,\ a \times 0 \neq 1$.

En particulier, l'ensemble $\mathbf{Q}^*$ *des seuls rationnels positifs* (zéro exclu ainsi que tous les négatifs) est un groupe abélien par rapport à la multiplication : l'inverse d'un rationnel *positif* est un rationnel *positif* ; le produit de deux rationnels positifs est un rationnel positif (contrairement à ce qui se passe avec les négatifs : le produit de deux négatifs est positif).

b) *Un ordre (de type η)*

Comme les entiers, les rationnels peuvent être *ordonnés*. On rappelle qu'un rationnel p/q est dit plus grand qu'un autre rationnel p'/q' si la différence $p/q - p'/q'$ est positive :

$$p/q > p'/q' \Leftrightarrow p/q - p'/q' > 0$$

Exemple :

$$2/3 > 3/5 \text{ car } 2/3 - 3/5 = \frac{10-9}{15} = \frac{1}{15} > 0.$$

$(\mathbf{Q}, >)$ est un ensemble totalement ordonné ; mais, entre deux rationnels, on *peut toujours en insérer un troisième* ; en effet, on a par exemple :

$$0 < p/q < p'/q' \Rightarrow p/q < \frac{p+p'}{q+q'} < p'/q'$$

Il n'y a donc pas de rationnel qui suive un rationnel donné.

Exemple :

Après 1/2, il y a 2/3, $1/2 < 2/3$; mais entre 1/2 et 2/3, il y a 3/5 : $1/2 < 3/5 < 2/3$.

Entre 1/2 et 3/5, il y a 4/7 : $1/2 < 4/7 < 3/5$.

Entre 1/2 et 4/7, il y a 5/9 : $1/2 < 5/9 < 4/7$, etc.

Exercice :

Prouver que $0 < p/q < p'/q' \Rightarrow p/q < \dfrac{p+p'}{q+q'} < p'/q'$.

Puisqu'entre deux rationnels on peut toujours en insérer un troisième (distinct des deux premiers), on peut donc en insérer une *infinité* : entre deux rationnels quelconques, il y a *toujours* une *infinité d'intermédiaires.*

G. Cantor désignait le type d'ordre de $\mathbf{Q}$ par la lettre η *(êta)* : le type η (cf. § 7). On voit que ce type, qui permet toujours l'insertion d'intermédiaires, est radicalement différent des types *discrets* ω et ζ.

$(\mathbf{Q}, >)$ est un ensemble totalement ordonné tel qu'entre deux éléments quelconques il y ait *toujours au moins un intermédiaire,* et par suite une *infinité d'intermédiaires.* Il en est de même de $(\mathbf{Q}^*, >)$.

c) *Groupes ordonnés*

$(\mathbf{Q}, +, >)$ est un groupe abélien ordonné : l'addition des rationnels est compatible avec l'ordre, dans le sens qui a été donné (§ 2 *c*) à la compatibilité.

$(\mathbf{Q}, +, >)$ est un groupe abélien ordonné

De même, $(\mathbf{Q}^*, \times, >)$ est un groupe abélien ordonné ; puisque tous les nombres qu'il contient sont *positifs,* on a bien la propriété :

$$x > y \Rightarrow x \times z > y \times z$$

$(\mathbf{Q}^*, \times, >)$ est un groupe abélien ordonné

d) *Un corps*

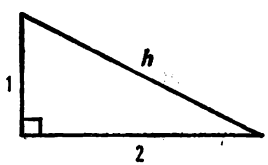
Muni de l'addition et la multiplication, **Q** est un *corps*. Un *corps* est un ensemble dans lequel sont définies *deux* opérations, qui est un groupe abélien pour la première, qui, lorsqu'on en exclut l'élément neutre de la première, est un groupe pour la seconde, et tel que la seconde opération soit distributive par rapport à la première. Comme la multiplication est commutative, on dit que **Q** est un *corps commutatif*; à cause des propriétés de l'ordre, que c'est un *corps ordonné*.

$(\mathbf{Q}, +, \times)$ est un corps commutatif

$(\mathbf{Q}, +, \times, >)$ est un corps ordonné

5. Les réels : R et R*

Il est bien connu que les nombres rationnels ne permettent pas la mesure de toutes les grandeurs que l'on peut être amené à considérer : par exemple, en géométrie, l'hypoténuse h du triangle rectangle de côtés 1 et 2 a pour carré 5 d'après la relation de Pythagore :



$$h^2 = 1^2 + 2^2 = 1 + 4 = 5$$

Mais aucun nombre rationnel n'est de carré égal à 5. Tout au plus pourra-t-on définir des rationnels dont le carré est *approximativement* égal à 5 (cf. § 6 et 8).

On introduit donc, pour la commodité des calculs *algébriques*, une nouvelle espèce de nombres, comprenant outre les

rationnels, tous les irrationnels tels que $\sqrt{5}, \sqrt{2}, \sqrt[7]{9}$, etc. Cela permet de noter, et de faire entrer dans des calculs littéraux ou algébriques, des nombres pour lesquels le calcul *numérique* n'est possible que de façon *approximative* (cf. § 7). Ce qui justifie l'introduction de ces nombres, c'est aussi la commodité du calcul exponentiel et logarithmique, étudié dans le prochain chapitre. Ces nombres sont (assez paradoxalement) appelés *réels*, et leur ensemble est noté **R**.

Les nombres réels, comme les rationnels, peuvent être structurés en :

— groupe abélien ordonné par rapport à $+$ et $>$:

$$(\mathbf{R}, +, >)$$

— corps commutatif ordonné par rapport à $+$, $\times$, et $>$:

$$(\mathbf{R}, +, \times, >)$$

— les réels *positifs* $\mathbf{R}^*$ (zéro exclu) constituent un groupe abélien ordonné par rapport à $\times$ et $>$:

$$(\mathbf{R}^*, \times, +)$$

L'ordre des nombres réels est évidemment tel qu'entre deux quelconques d'entre eux il y en ait toujours un troisième, donc une infinité.

6. Les décimaux : D

On démontre qu'il n'existe aucun système d'écriture, tel qu'un système de numération, par exemple, qui permette d'écrire *tous* les nombres réels, alors que l'on dispose pour les entiers et les rationnels de tels systèmes d'écriture, qui sont bien connus. On écrit tous les nombres entiers avec un « alphabet » de 11 signes, dans le système dit de numération

décimale ; ces 11 signes sont les 10 chiffres 0, 1, ..., 9, et le signe « — » qui indique les nombres négatifs :

$$1\,967, \quad -32, \quad \text{etc.}$$

Avec un signe supplémentaire, la barre de fraction « / », on écrit tous les rationnels :

$$-2/5, \quad 272/395, \quad \text{etc.}$$

En ce qui concerne *les réels*, on ne pourra en général qu'en *écrire des approximations* ; ces approximations peuvent se faire au moyen des rationnels : 5/3, 17/10, etc., sont des approximations de $\sqrt{2}$.

Mais on utilise plus couramment comme système d'écriture l'ensemble **D** des nombres décimaux ; leur écriture se fait au moyen de 12 signes (les dix chiffres ; « — » ; la virgule « , »).

Par exemple, on approche π par 3 ; puis 3,1 ; puis 3,14 ; puis 3,142 ; puis 3,1416 ; etc.

L'écriture des nombres décimaux est très commode car il suffit de *lire* deux nombres décimaux pour pouvoir les *comparer*, savoir dans quel ordre ils sont ; la lecture suffit pour savoir que $3{,}142 > 3{,}1416$. Alors que l'écriture des fractions est impropre à la comparaison à vue : quelle est la plus grande des deux fractions 272/395 et 3 596/4 612 ?

Les décimaux sont également plus commodes pour effectuer des additions et des multiplications.

Par contre, les nombres décimaux se prêtent moins bien au calcul par division que les rationnels : l'inverse d'un décimal n'est en général pas un décimal. Exemple : l'inverse de 1,4 est 5/7 qui n'est pas un nombre décimal (seules les fractions dont le dénominateur est de la forme $2^a \times 5^b$ sont de nombres décimaux : pourquoi ?). **D**, par rapport à l'addition et à la multiplication, ne constitue qu'un *anneau*, ce n'est pas un corps. L'ordre des décimaux est, comme celui

de **Q** et **R**, tel qu'entre deux décimaux on peut toujours en insérer un troisième ; entre 3,14 et 3,15 il y a 3,142 par exemple ; on obtient l'insertion *d'éléments intermédiaires par simple* adjonction de décimales supplémentaires.

Dans la pratique du calcul numérique, on se sert de l'une ou l'autre écriture (fractions, décimaux) selon l'usage que l'on en fait : on adopte l'écriture la mieux adaptée aux calculs que l'on doit effectuer ; en cas d'ambiguïté (quand on doit faire par exemple à la fois des divisions et des comparaisons), on se sert simultanément des deux systèmes. Le tableau suivant résume ces avantages des deux systèmes d'écriture approximative des nombres réels.

	Rationnels **Q**	Décimaux **D**
Addition +		*
Multiplication ×		*
Division :	*	
Comparaison >		*

Note :

1. Le système d'écriture décimal n'est pas le seul de son espèce ; il se généralise naturellement à tout système dans lequel on adopte un alphabet de 2 chiffres au moins, avec un ordre sur les chiffres, et complété par les signes « — » et « , ». Par exemple, dans le système ternaire, l'alphabet a 3 chiffres, souvent notés 0, 1 et 2, dans l'ordre $0 < 1 < 2$. Le rôle joué dans le système décimal par les puissances de 10 est tenu ici par les puissances de 3.

2. L'importance capitale des nombres décimaux pour la mesure tient au fait qu'ils constituent un système très pratique d'écriture pour les calculs numériques, et pour

la comparaison ; en particulier la plupart des appareils de mesure (de longueur, de temps, de température, etc.) sont gradués en nombres décimaux ; ils fournissent par conséquent une valeur approximative, en nombres décimaux, de la mesure que l'on veut repérer.

*7. Types d'ordre (échelles) : la comparaison

Les diverses espèces de nombres que l'on a rencontrées sont en particulier munies d'une relation d'ordre ; on a vu que l'ordre n'a pas les mêmes propriétés dans **N** et dans **Q**, par exemple. Si l'on considère un ensemble totalement ordonné, ayant pour son ordre les propriétés de celui de **N**, mais où ne sont pas définies d'*opérations* telles que l'addition ou la multiplication, on dit qu'on a affaire à un ordre de *type* ω *(omega)* :

> *Ordre de type* ω : ensemble infini totalement ordonné, qui possède un plus petit élément, et tel qu'entre deux éléments il y ait toujours un *nombre fini d'intermédiaires.*

Exemple :

La suite ordonnée des entiers ordinaux :

$$1^{er} < 2^{e} < 3^{e} < 4^{e} < \dots$$

De même, tout ensemble totalement ordonné est dit de *type* ζ *(dzeta)* si son ordre total a les propriétés de celui de **Z.**

> *Ordre de type* ζ : ensemble infini totalement ordonné, n'ayant pas de plus petit élément, et tel qu'entre deux éléments quelconques il y ait toujours un nombre *fini d'intermédiaires.*

Exemple :

Le repérage des dates :

$\ldots < 2$ avant < 1 avant $< 0 < 1$ après < 2 après $< \ldots$

Les ensembles totalement ordonnés *finis*, dont l'ordre est dit de *type fini*, ont évidemment comme ω et ζ la propriété qu'entre deux éléments, il n'y a jamais qu'un nombre fini d'intermédiaires ; c'est cette propriété que l'on traduit en disant que ω, ζ et les types finis sont des *types discrets*.

Un autre type d'ordre usuel est celui qui est désigné par η dans les notations de G. Cantor ; la propriété 1 de sa définition le distingue radicalement des types discrets.

Ordre de type η : Un ensemble totalement ordonné est dit de *type* η *(êta)* si, comme $(\mathbf{Q}, >)$ ou $(\mathbf{D}, >)$, il est tel :

1) qu'on puisse toujours, entre deux éléments, en insérer un troisième (donc une infinité) ;
2) qu'on puisse noter *tous* ses éléments, avec un alphabet *fini*, et par des « mots » de *longueur finie* ;
3) il n'a ni plus petit élément, ni plus grand élément.

La condition 2 s'exprime souvent en disant que l'ensemble est *dénombrable* : ceci signifie qu'on peut *nommer* chacun de ses éléments en n'écrivant que des « mots » de longueur finie au moyen de « lettres » prises dans un alphabet fini. Cette condition est nécessaire pour que soit accessible aux moyens humains de communication — langage ou écriture — un ensemble infini.

Les types ω, ζ et η sont tous trois dénombrables ; il y a des ensembles totalement ordonnés qui sont dénombrables, et n'appartiennent à aucun de ces types ; par exemple :

$$1 < 2 < 3 \ldots ; \quad \mathrm{I} < \mathrm{II} < \mathrm{III} < \mathrm{IV} \ldots$$

avec la convention que tous les numéros écrits en chiffres arabes sont avant tous les numéros écrits en chiffres romains. Entre deux numéros « arabes » il n'y a qu'un nombre fini d'intermédiaires ; de même entre deux numéros « romains ». Mais entre un « arabe » et un « romain », il y en a une infinité.

Il y a d'autre part des ensembles totalement ordonnés qui ne sont pas dénombrables ; en particulier c'est le cas de $(\mathbf{R}, >)$.

En effet, on ne pourrait noter tous les éléments de $(\mathbf{R}, >)$ avec un alphabet fini qu'en « écrivant » des mots de longueur infinie, c'est-à-dire comportant une infinité de lettres (cf. § 5 et 8) ; une telle « écriture » ne sera jamais effectivement réalisée.

Dans les sciences sociales, où les ensembles totalement ordonnés sont souvent utilisés sous le nom d'*échelles*, parfois sous celui d'*échelles ordinales* (échelles d'attitude, de sensation, de préférence...), les types ω, ζ et η suffisent à rendre compte de la plupart des situations étudiées ; c'est en particulier à la propriété 1 des échelles η (entre deux éléments, il y a toujours un intermédiaire) que l'on fait en général allusion lorsqu'on parle d' « échelles continues » ou de « continuum ».

Le mot *continu* a plusieurs sens en mathématiques ; dans la typologie de l'ordre, on appelle parfois ordre *continu* un ordre total qui, comme $(\mathbf{R}, >)$, est approximable par un ordre de type η ; ceci n'est pas la définition précise (qui ne sera pas donnée ici), mais fournit simplement un exemple d'ordre continu.

N.B. — 1. Deux ensembles totalement ordonnés de même type sont bien équivalents en ce qui concerne leur structure d'ordre ; on démontre en effet que si deux ensembles totalement ordonnés sont par exemple de type η (satisfont aux trois propriétés encadrées de la définition de η) on peut

établir de l'un à l'autre une bijection monotone (*i.e.* conservant l'ordre).

2. On a dit que le mot *échelle* désigne en général un ordre total dans la méthodologie des sciences humaines ; celle-ci distingue parfois :

— échelle *nominale*, ce qui est synonyme d'*ensemble* ;
— échelle *qualitative* ou *ordinale*, ce qui est synonyme d'*ordre total* ;
— échelle *quantitative* ou *numérique*, ce qui est en général synonyme de groupe abélien totalement ordonné.

8. Approximations

a) *Notion courante, notion précise*

Quand on demande 1 kg de viande au boucher, on sait qu'il vend à peu près 1 kg, mais pas exactement 1 kg ; ce serait d'ailleurs impossible pour des raisons évidentes : précision relative de la balance, par exemple.

On a donc une certaine notion de l' « à-peu-près » ; on va essayer de la préciser. Cette précision est donnée par l'ajusteur qui dit faire un cylindre de 3,5 cm de rayon à 2 microns près, ce qui veut dire que le « vrai » diamètre de la pièce, inconnu et impossible à connaître, est compris entre (3,5 — 0,002) mm et (3,5 + 0,002) mm : on a ici une notion précise de l' « à-peu-près ». D'où la définition :

> Soit x un nombre ; on dit que x' est une *valeur approchée* de x à ε près si :
>
> $$x' - \varepsilon < x < x' + \varepsilon$$
>
> ε est appelé l'*erreur* que l'on commet sur x.

On peut écrire : $x \simeq x'$ (à ε près).
Le signe $\simeq$ se lit : « est approximativement égal à ».

b) *Calcul de l'approximation d'un nombre réel par un nombre décimal*

Si l'on veut calculer la surface d'un cercle dont on connaît le rayon r, on sait que : $S = \pi r^2$.

Il faut donc connaître π ; π est un nombre réel irrationnel (représentant le rapport de la longueur de la circonférence à celle du diamètre).

Pour faire le calcul numérique, il faudrait une écriture au moyen de chiffres du nombre π. On a vu (§ 5 et 6) que celle-ci est impossible. Plus précisément, cette écriture nécessiterait une infinité de chiffres ; comme on ne peut jamais effectivement écrire que des chiffres en nombre fini (actuellement, on connaît environ 500 000 décimales du nombre π) ce sont seulement des approximations de π que l'on écrit. Suivant les besoins, on dira que :

π est à peu près égal à 3
— — 3,14
— — 3,1416
— — 3,14159265358

— Tous les nombres qui ont été écrits, sont des nombres décimaux et l'on dit qu'ils sont de plus en plus précis ; mais qu'est-ce que cela signifie ? Autrement dit, quelle est l'erreur faite en remplaçant π par l'un de ces nombres ?

— Par convention, dire que π est à peu près égal à 3,14, c'est dire que le nombre π est compris entre deux bornes :

$$3{,}14 - 0{,}005 < \pi < 3{,}14 + 0{,}005$$
$$3{,}135 < \pi < 3{,}145$$

3,135 et 3,145 sont des nombres décimaux exacts.

L'erreur est au plus de $0{,}5 \times 1/100$.

De même dire que la valeur de π est à peu près 3,1416, c'est dire que le nombre π est compris entre deux bornes :

$$3{,}1416 - 0{,}00005 < \pi < 3{,}1416 + 0{,}00005$$
$$3{,}14155 < \pi < 3{,}141665$$

L'erreur est donc ici de $0{,}5 \times 1/10\,000$ au plus.

Plus généralement, si l'on donne une valeur approchée d'un nombre avec n décimales exactes, l'erreur est au plus de :

$$0{,}5 \times \frac{1}{\underbrace{10 \times 10 \times \ldots \times 10}_{n \text{ fois}}}$$

ce qui se note $0{,}5/10^n$.

— Cette convention est très importante et naturelle ; son application permet d'éviter les précisions illusoires.

Si, par exemple, un nombre réel admet comme valeur approchée 21,6234, avec une erreur de 1/100 d'unité, la précision des deux derniers chiffres 3 et 4 ne signifie rien.

On a en effet :

$$21{,}6234 - 1/100 < a < 21{,}6234 + 1/100$$

ou :

$$21{,}6134 < a < 21{,}6334$$

On voit que les deux dernières décimales 3 et 4 n'ont aucun sens, et qu'avec à peu près la même erreur on peut écrire 21,6229. Puisque ces deux derniers chiffres ne servent à rien, n'ont aucune signification, il est inutile de les écrire et on dira que la valeur approchée de a est 21,62 à 1/100 près ou que a est donné avec 2 décimales exactes.

c) *Faire des opérations avec des nombres approchés*

Quand on remplace un nombre par sa valeur décimale approchée, c'est en général pour faire des calculs numériques. Soit, par exemple, une entreprise qui reçoit de diffé-

rents fournisseurs des quantités en kilogrammes d'un certain produit :

1re livraison :	3,45	kg	à peu près ;	l'erreur est	5	g ;
2e — :	3,25	kg	—	—	5	g ;
3e — :	10,20	kg	—	—	5	g ;
4e — :	2,575	kg	—	—	5	dg.

Le responsable du stock fait son addition et dit qu'il a en magasin à peu près 19,475 kg. Mais tous les termes entrant dans la somme ne sont que des valeurs approchées ; aussi faut-il connaître l'erreur faite sur la somme. C'est, rappelons-le, un nombre ε tel que l'on soit sûr que la vraie valeur p, en kilogrammes, du stock est comprise entre :

$$19{,}475 - \varepsilon \text{ et } 19{,}475 + \varepsilon$$

Sur chacun des termes, l'erreur a pu être faite soit par excès, soit par défaut, au pire, elle a été faite par excès sur *tous* les termes, ou par défaut sur tous les termes. La seule valeur que l'on puisse par conséquent assigner à ε est, en grammes :

$$\varepsilon = 5 + 5 + 5 + 0{,}5 = 15{,}5$$

Soit 0,0155 kg. Le poids total est compris entre $(19{,}475 - 0{,}0155)$ kg et $(19{,}475 + 0{,}0155)$ kg :

$$19{,}4595 < p < 19{,}4905$$

On voit que le dernier chiffre de la valeur 19,475 est illusoire, et que le poids est à peu près de 19,47 kg, à 15 g près.

— Des formules permettent de déterminer l'erreur faite dans un calcul où interviennent plusieurs opérations.

Si on appelle $x, y, z, \ldots$ les nombres sur lesquels on opère et $\Delta x, \Delta y, \Delta z, \ldots$ les erreurs correspondantes, on calcule par les égalités suivantes l'erreur sur le résultat d'une opération.

Erreur sur la somme :

$$\text{Si } S = x + y + z \quad \text{alors } \Delta S = \Delta x + \Delta y + \Delta z$$

Erreur sur la différence (attention, ce n'est pas la différence des erreurs, mais leur somme) :

$$\text{Si } S = x - y \quad \text{alors } \Delta S = \Delta x + \Delta y$$

Erreur sur le produit (elle se calcule au moyen de l'erreur *relative*) :

$$\text{Si } P = xyz \quad \text{alors } \frac{\Delta P}{P} = \frac{\Delta x}{x} + \frac{\Delta y}{y} + \frac{\Delta z}{z}$$

Erreur sur le quotient :

$$\text{Si } Q = \frac{x}{y} \quad \text{alors } \frac{\Delta Q}{Q} = \frac{\Delta x}{x} + \frac{\Delta y}{y}$$

Exemples :

1. $S = x - y$.

On sait que :

$$x = 50 \quad \Delta x = 0{,}1$$
$$y = 20 \quad \Delta y = 0{,}2$$
$$S = 50 - 20 = 30 \text{ et } \Delta S = 0{,}1 + 0{,}2 = 0{,}3$$

2. $P = \frac{xy}{z}$ $\quad \frac{\Delta P}{P} = \frac{\Delta x}{x} + \frac{\Delta y}{y} + \frac{\Delta z}{z}$.

On sait que :

$$x = 50 \quad \Delta x = 0{,}1$$
$$y = 20 \quad \Delta y = 0{,}2$$
$$z = 10 \quad \Delta z = 0{,}1$$

Alors :

$$P = \frac{50 \times 20}{10} = 100$$

$$\frac{\Delta P}{P} = \frac{0{,}1}{50} + \frac{0{,}2}{20} + \frac{0{,}1}{10} = \frac{22}{1\,000}$$

D'où :

$$\Delta P = 100 \times \frac{22}{1\,000} = 2{,}2$$

9. De quels nombres faut-il se servir ?

Des diverses espèces de nombres définies dans ce chapitre, lesquels utilise-t-on, et pour quoi faire ? Il faut ici distinguer trois phases dans l'utilisation de nombres :

1) l'expression numérique des données expérimentales ou des observations ;
2) le modèle par lequel on interprète ces données, et les calculs *algébriques* auxquels celui-ci conduit ;
3) les calculs *numériques* sur les expressions algébriques obtenues à partir du modèle.

1. Les *données*, les observations, si elles ne sont pas simplement qualitatives, mais lorsqu'elles sont *numériques*, sont :

— soit une série de mesures effectuées sur des « variables », et sont généralement exprimées en *nombres décimaux*, ou en *entiers* positifs ou négatifs ;
— soit des dénombrements (une statistique), et sont alors exprimées en nombres *entiers naturels*.

Les *données* : entiers **N, Z** décimaux **D, D***

2. Pour le *modèle*, les lois théoriques qui sont supposées régir les phénomènes étudiés, on se donne le maximum de facilités de calcul *algébrique*. On utilise donc la structure numérique la mieux adaptée aux calculs que l'on a en vue, pourvu qu'elle constitue une bonne *approximation* des données empiriques.

Si ces calculs n'impliquent que des additions, des soustractions et des multiplications, alors il est recommandé de rester dans la structure par laquelle s'expriment les données, **Z** ou **D**, puisque tous ces calculs sont possibles dans la structure d'*anneau*.

Si l'on veut faire des divisions, des changements d'unité, et plus généralement avoir toutes les possibilités de l'*algèbre linéaire*, il faut une structure de corps ; on se place donc dans **Q**, le *corps des rationnels*, dont on a vu par ailleurs qu'il constitue une bonne approximation de **D** (mais pas de **Z** : si les données sont en nombres entiers, faire les calculs algébriques, au niveau du modèle, en nombres rationnels, nécessite des précautions, et une critique des résultats obtenus).

Si l'on veut enfin pouvoir calculer des racines carrées (calculs de distances, de corrélations, chap. IV et V), ou plus généralement faire du *calcul exponentiel et logarithmique* (chap. II), alors il faut utiliser le *corps* **R** *des réels*, dans lequel ces opérations sont définies, alors qu'elles ne le sont pas dans **Q**. D'autre part, **R** constitue lui aussi une bonne approximation de **D** et de **Q** (mais non de **Z**).

Les calculs *algébriques* (modèle) :
Toutes les structures numériques. Mais :
Q si l'on fait du *calcul linéaire* ;
R si l'on fait du *calcul exponentiel* ou *logarithmique* (en particulier, racines carrées).

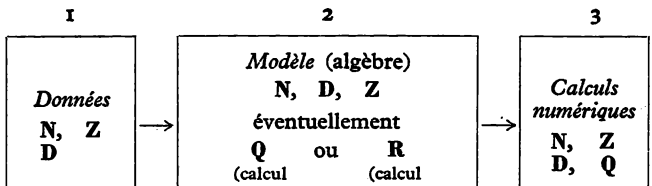
3. Pour les *calculs numériques* sur les expressions algébriques résultant du modèle, force est bien de revenir à une

structure numérique dans laquelle tous les nombres puissent s'écrire, se noter, c'est-à-dire **D**, **Q**, ou **Z** (ou **N**). On choisit là encore la mieux adaptée aux calculs à effectuer (cf. § 5).

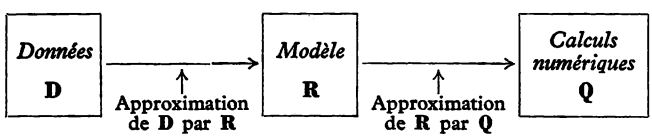
Si les calculs algébriques ont été faits par exemple dans **R**, comme **D** et **Q** sont de bonnes approximations de **R**, on peut, comme cela est indiqué au § 8, calculer l'erreur commise dans cette substitution de **D** (ou **Q**) à **R**.

Calculs *numériques* : **N, Z**
D, Q

En résumé :



Et retenons qu'il y a *deux fois* une *approximation* ; soit par exemple le schéma :



Il y a une première approximation quand on remplace la structure **D** dans laquelle s'expriment les données par **R** ; et une seconde quand on remplace **R** par la structure **Q** dans laquelle sont faits les calculs numériques.

Récapitulation

Calculer — Structures numériques fondamentales	*Comparer* — Types d'ordre	
N : entiers naturels 0, 1, 2, ... ; $+, \times$	ω	discrets ; dénombrables
Z : *anneau* des entiers relatifs ; $+, -, \times$	ζ	discrets ; dénombrables
D : *anneau* des nombres décimaux ; $+, -, \times$	η	dénombrables
Q : *corps* des rationnels ; $+, -, \times, :$	η	dénombrables
R : *corps* des réels ; $+, -, \times, :$		non dénombrable

Les réels, **R**, servent pour des *calculs algébriques* (en particulier : exponentiels et logarithmiques).

Le *calcul numérique* sur les réels se fait par *approximations*, avec les nombres rationnels **Q**, ou décimaux **D**.

CHAPITRE II

FONCTIONS NUMÉRIQUES

Les fonctions numériques étudiées dans ce chapitre sont des applications dont l'ensemble de départ et l'ensemble d'arrivée sont des ensembles de nombres. Toutes sont liées aux structures, étudiées au chapitre I, des ensembles de nombres.

1. LA FONCTION LINÉAIRE

a) *Fonction linéaire de* D (N, Z, Q)

Quand un négociant achète ou vend des pommes de terre, à une certaine quantité x de ce produit représentée par un nombre, on fait correspondre un prix qui ne dépend (à une date donnée) que de la quantité x ; il sera noté $f(x)$. La correspondance $x \to f(x)$ est une application $f : \mathbf{D} \to \mathbf{D}$; quantités et prix sont en effet généralement exprimés en nombres décimaux.

Quelles sont les propriétés de cette application ?

Si le négociant achète une quantité de x, il paie $f(x)$.

S'il achète la quantité y, il paie $f(y)$.

S'il achète la quantité $(x + y)$, il paie $f(x + y)$.

En général le prix payé pour $x + y$ est la somme des prix payés pour x et y ; on peut donc écrire, s'il en est ainsi :

$$f(x + y) = f(x) + f(y)$$

Cette application a donc la propriété de *conserver les additions :* l'image d'une somme de deux nombres est la somme des images de ces nombres. Toute application de **D** dans **D** ayant cette propriété est appelée *fonction linéaire de* **D.**

Si, dans cette définition, l'ensemble **D** des nombres décimaux est remplacé par **N**, ou **Z**, ou **Q**, on dit qu'on a une fonction linéaire de **N** dans le premier cas, de **Z** dans le second, de **Q** dans le troisième.

Fonction linéaire de **D** (ou **N**, ou **Z**, ou **Q**) : toute application $f : \mathbf{D} \to \mathbf{D}$ (resp. **N**, **Z**, **Q**) telle que *l'image d'une somme* de deux nombres soit la *somme des images* de ces nombres.

$$\forall x \in \mathbf{D}, \quad \forall y \in \mathbf{D}, \quad f(x+y) = f(x) + f(y)$$

b) *Formule et graphique d'une fonction linéaire*

— On démontre que toute fonction linéaire peut se mettre sous la forme :

$$f(x) = a \times x$$

où a est une constante, le *coefficient* de la fonction linéaire.

$$f(1) = a$$

a est l'image de 1

Dans l'exemple, a est le prix du kilogramme de pommes de terre.

— On vérifie d'ailleurs immédiatement, d'après la *distributivité* de la multiplication par rapport à l'addition, que si : $f(x) = a \times x$, on a :

$$f(x+y) = a \times (x+y) = a \times x + a \times y = f(x) + f(y)$$

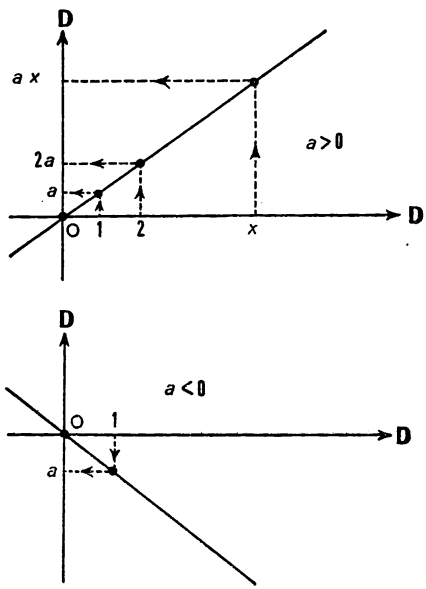
La réciproque est vraie ; si f est une application de **D** dans **D** telle que :

$$f(x+y) = f(x) + f(y)$$

alors : $f(x) = a \times x$.

Cette réciproque est délicate à démontrer (1).

— Si l'on représente dans un graphique cartésien une fonction linéaire, ce graphique est une *droite* de *pente a*, passant par l'origine des axes.



(1) En fait, on peut montrer que toute application f du groupe abélien $(\mathbf{Z}, +)$ par exemple dans lui-même conservant l'addition (endomorphisme) est déterminée par l'image a du nombre 1 dans cette application ; et poser par *définition* de la multiplication : $a \times x = fa(x)$, où fa est l'endomorphisme de $(\mathbf{Z}, +)$ tel que $fa(1) = a$.

c) *Fonction linéaire de* **R**

La formule $\boxed{f(x) = a \times x}$ reste bien définie lorsque x peut être irrationnel ; c'est-à-dire lorsque l'ensemble de départ de l'application f est l'ensemble **R**, et pas seulement **D** ou **Q**.

On continue d'appeler *fonction linéaire de* **R** toute application $f : \mathbf{R} \to \mathbf{R}$ telle que l'image de chaque nombre réel x soit donnée par la formule :

$$f(x) = a \times x \quad (a \text{ nombre réel donné})$$

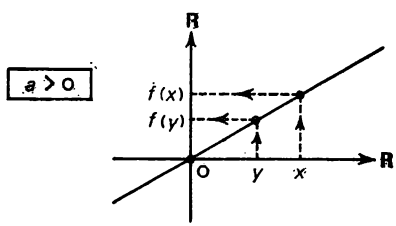
Le graphique de cette fonction est toujours une droite passant par l'origine et de pente a.

Mais, dans ce cas, la seule propriété de conserver les additions ne suffit plus à caractériser les fonctions linéaires, n'est plus équivalente à $f(x) = a \times x$.

Il faut une condition supplémentaire, qui est la *monotonie* de f.

Comme on le remarque sur les graphiques, si $a > 0$, la fonction est monotone *croissante* :

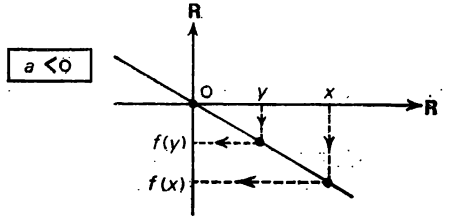
$$x > y \Rightarrow f(x) > f(y)$$



Autrement dit, dans ce cas, la fonction f *conserve l'ordre* de $(\mathbf{R}, >)$.

Si $a < 0$, la fonction est monotone *décroissante* :

$$x > y \Rightarrow f(x) < f(y)$$



Autrement dit, *f inverse l'ordre* de $(\mathbf{R}, >)$.

On démontre (on admet ici) qu'une fonction $f : \mathbf{R} \to \mathbf{R}$ est *linéaire* si, et seulement si :

1° L'image d'une somme est la somme des images :

$$\forall x, \quad \forall y \qquad f(x + y) = f(x) + f(y)$$

2° Elle *conserve* l'ordre :

$$\forall x, \quad \forall y \qquad x > y \Rightarrow f(x) > f(y)$$

ou

2° *bis*. Elle *inverse* l'ordre :

$$\forall x, \quad \forall y \qquad x > y \Rightarrow f(x) < f(y)$$

Cette définition *équivaut* à :

$$f(x) = a \times x$$

La fonction linéaire de coefficient $a \neq 0$ possède une fonction *inverse* f^{-1} (1) :

$$f^{-1}(x) = \frac{1}{a} \times x$$

(1) Pour la définition de l'inverse d'une bijection, et de la composition o des fonctions, cf. *MSH*, chap. II.

En effet :

$$f \circ f^{-1}(x) = a \times \left(\frac{1}{a} \times x\right) = x$$

$$f^{-1} \circ f(x) = \frac{1}{a} \times (a \times x) = x$$

Une fonction linéaire est une *bijection.*

d) *Automorphismes*

On a vu que $(\mathbf{R}, +, >)$, c'est-à-dire $\mathbf{R}$ muni de l'*addition* et de l'*ordre total* constitue un *groupe abélien* ordonné (cf. chap. I, § 5). Les fonctions linéaires de $\mathbf{R}$, qui sont *bijectives*, et *conservent* la structure puisqu'elles conservent les additions et l'ordre, sont des *automorphismes* de la structure de groupe abélien ordonné $(\mathbf{R}, +, >)$.

Dans ce terme, *morphisme* signifie *application conservant la structure* (*morphe :* forme), et *auto* signifie que cette application est d'un ensemble dans lui-même ($\mathbf{R}$ est départ et arrivée d'une fonction linéaire).

Ceci se symbolise (et se résume) par le diagramme :

$$(\mathbf{R}, +, >) \underset{\square \times \frac{1}{a}}{\overset{\square \times a}{\underset{\longleftarrow}{\overset{\longrightarrow}{\text{Lin.}}}}} (\mathbf{R}, +, >) \qquad a \neq 0$$

Exercices :

— Montrer que $f : \mathbf{Q} \to \mathbf{Q}$ définie par $f(x) = a \times x$, $a \in \mathbf{Q}$, est un automorphisme de $(\mathbf{Q}, +, >)$.

— Montrer que $f : \mathbf{Z} \to \mathbf{Z}$ définie par $f(x) = a \times x$, $a \in \mathbf{Z}$, conserve la structure $(\mathbf{Z}, +, >)$ mais n'est pas bijective (sauf si $a = 1$). C'est un *endomorphisme.*

— Montrer que si $f : \mathbf{R} \to \mathbf{R}$ a la propriété : $\forall \lambda \in \mathbf{R}$, $\forall x \in \mathbf{R}$, $f(\lambda x) = \lambda f(x)$, alors f est linéaire.

I *bis*. FONCTIONS AFFINES

a)

On appelle *fonction affine* de **R** (respectivement **Q**, **D**, **Z**) toute application $f : \mathbf{R} \to \mathbf{R}$ (respectivement **Q**, **D**, **Z**) telle que :

$$\forall x \in \mathbf{R} \qquad f(x) = a \times x + b$$

a et b sont des constantes données, appartenant à l'ensemble de nombres sur lequel f est définie ; a est le *coefficient* de la fonction affine, b son *ordonnée à l'origine.*

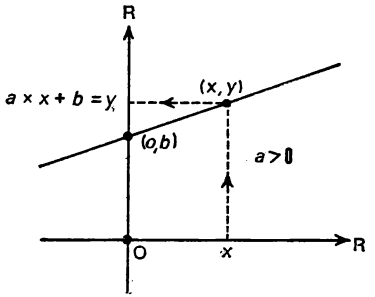
Cette appellation de la constante b se comprend d'après la représentation graphique de f, qui est une *droite.* Lorsque $x = 0$ (point à l'origine), on a :

$$f(x) = f(0) = a \times 0 + b = b$$

Donc l'ordonnée correspondant à l'origine vaut b, la droite coupe l'axe vertical (des ordonnées) au point $(0, b)$.

La fonction affine est, comme la fonction linéaire :

— monotone croissante lorsque $a > 0$;
— monotone décroissante lorsque $a < 0$.



On démontre que la propriété caractéristique des fonctions affines est, outre de conserver (ou inverser) l'ordre, de *conserver les différences* :

$$\boxed{f \text{ est affine}} \Leftrightarrow \boxed{x - y = z - t \Leftrightarrow f(x) - f(y) = f(z) - f(t)}$$

Ce sont donc les fonctions affines qui sont associées à ce que la psychologie expérimentale nomme les échelles d'*intervalles* ; une échelle d'intervalles est définie à une fonction affine près.

Exercice :

Prouver la propriété ci-dessus :

a) Montrer que si $f(x) = a \times x + b$, on a :

$$x - y = z - t \Leftrightarrow f(x) - f(y) = f(z) - f(t)$$

b) Énoncer la réciproque de *a)*.

c) Démontrer cette réciproque ; poser :

$$f(x) - f(0) = h(x) \qquad f(0) = b$$

et montrer que h est *linéaire*.

b) *Fonctions affines de* N *et* Z *: progressions arithmétiques*

Si f, donnée par la formule :

$$f(x) = a \times x + b$$

est une fonction en nombres *entiers*, par exemple *naturels*, f est une application de N dans N, si le coefficient a est entier, de même que b.

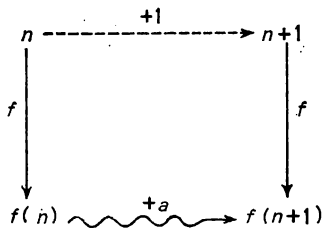
Dans ce cas, on a :

$$f(0) = b, \qquad f(1) = a + b = a + f(0),$$
$$f(2) = 2a + b = a + f(1)$$

Et d'une façon générale :

$$f(n+1) = a(n+1) + b = a + a \times n + b = a + f(n)$$

L'image $f(n+1)$ de l'entier $(n+1)$ *suivant* n se déduit de $f(n)$ en ajoutant le nombre constant a :



La suite des nombres $f(0), f(1), f(2), \ldots$ constitue une *progression arithmétique* de raison a. Réciproquement, toute progression arithmétique doit être considérée comme une *fonction affine* de **N** ou **Z**.

2. La fonction exponentielle

a) *Transformation de sommes en produits*

Dans le rapport du V^e^ Plan, on lit : « Le produit national brut de la France doit augmenter de 5 % par an. » C'est-à-dire que si, dans une certaine unité, ce produit est 1 au 1^er^ janvier de l'année d'application 1966, il sera de $1 + 0{,}05 = 1{,}05$ au 1^er^ janvier de l'année 1967. Que sera-t-il au 1^er^ janvier 1968 ? 1,05 + 5 % d'augmentation donc :

$$1{,}05 + 1{,}05 \times 0{,}05 = 1{,}05 \times (1 + 0{,}05) = (1{,}05)^2$$

Plus généralement on voit que la suite des valeurs obtenues est donnée par le tableau suivant.

Année	Produit
0	1
1	1,05
2	$1{,}05 \times 1{,}05 = (1{,}05)^2$
3	$1{,}05 \times 1{,}05 \times 1{,}05 = (1{,}05)^3$
⋮	⋮
n	$(1{,}05)^n$

Ces valeurs forment une *progression géométrique* de raison 1,05 ; ici, ces valeurs sont toutes des nombres décimaux. Si, à chaque année n, on fait correspondre le produit national brut de cette année, soit $(1{,}05)^n$, on définit une application $f : \mathbf{N} \to \mathbf{D}$.

$$f(n) = (1{,}05)^n$$

Quelle est la propriété fondamentale de cette application ?

— à l'année n :

$$f(n) = (1{,}05)^n$$

— à l'année $(n + n')$:

$$f(n + n') = (1{,}05)^{n+n'}$$
$$= \underbrace{(1{,}05) \times \ldots \times (1{,}05)}_{n + n' \text{ fois}}$$

Ce qui peut s'écrire :

$$\underbrace{(1{,}05) \times \ldots \times (1{,}05)}_{n + n' \text{ fois}}$$
$$= \underbrace{(1{,}05) \times \ldots \times (1{,}05)}_{n \text{ fois}} \times \underbrace{(1{,}05) \times (1{,}05)}_{n' \text{ fois}}$$

$$(1{,}05)^{n+n'} = (1{,}05)^n \times (1{,}05)^{n'}$$

On a donc :

$$f(n+n') = f(n) \times f(n')$$

Ce qui s'énonce : l'image d'une *somme* de deux nombres est le *produit* des images de ces nombres.

Une telle fonction est dite *exponentielle.*

> La fonction *exponentielle* transforme l'addition en multiplication.

Elle est de la forme :

$$f(x) = a^x$$

Le nombre fixe a est appelé *base* de l'exponentielle. Dans l'exemple, la base est $a = 1{,}05$. Cette base est l'*image de 1 :*

$$f(1) = a^1 = a$$

b) *Conservation (ou inversion) de l'ordre*

Dans l'exemple, on remarque qu'en outre les valeurs prises : 1, 1,05, $(1{,}05)^2$, ..., etc., vont *en croissant* : la fonction exponentielle de base 1,05 est *monotone croissante* ; il en est ainsi chaque fois que la base a est supérieure à 1.

> Toute fonction exponentielle de *base* $a > 1$ est *monotone croissante :* $x > y \Rightarrow f(x) > f(y)$.

Si l'on prend la base $a = 1/2$, la suite des images est :

n	0	1	2	3	...	n	...
$(1/2)^n$	1	1/2	1/4	1/8	...	$1/2^n$	...

Ces valeurs vont en décroissant ; il en est ainsi chaque fois que la base a est comprise entre 0 et 1.

Toute fonction exponentielle de base a telle que :

$$0 < a < 1$$

est *monotone décroissante* : $x > y \Rightarrow f(x) < f(y)$.

Si l'on prenait une base négative, par exemple $a = -2$, il n'y aurait plus monotonie ; les valeurs prises oscillent.

n	0	1	2	3	...
$(-2)^n$	1	-2	4	-8	...

Exercice :

Construisez sur un même papier millimétré les graphiques des exponentielles de base 2, 1/2 et (-2).

Pour n'avoir à considérer que des fonctions qui, comme les fonctions linéaires, *conservent l'ordre* des nombres (ou *inversent* cet ordre), on réserve le nom de *fonctions exponentielles* aux seules fonctions de la forme $f(x) = a^x$ ayant une *base* a *positive.*

Exercice :

Montrer que la fonction exponentielle de base $a = 1$ est une *application constante.*

c) *Exposant zéro, exposants négatifs*

Jusqu'ici, la fonction exponentielle ayant pour base un nombre a décimal positif est considérée comme une application $f : \mathbf{N} \to \mathbf{D}$; son ensemble de départ est celui des *entiers naturels.*

Si la base est un rationnel positif, par exemple $a = 6/7$, l'exponentielle a pour ensemble d'arrivée l'ensemble **Q** des rationnels : $(6/7)^n$ est un rationnel quel que soit l'exposant entier naturel n. C'est une application $f : \mathbf{N} \to \mathbf{Q}$.

Ses propriétés fondamentales sont que :

1° l'image d'une somme est le produit des images ;
2° l'ordre est conservé (ou inversé).

— C'est la première de ces propriétés qui amène à poser :

$$\boxed{a^0 = 1}$$

En effet :

$$f(x + 0) = f(x) \times f(0)$$

Mais :

$$f(x + 0) = f(x) \qquad \text{puisque} \quad x + 0 = x$$

Donc, si $f(x) \neq 0$:

$$\boxed{f(0) = 1}$$

Soit, en posant $f(x) = a^x$: $a^0 = 1$.

— Cette même propriété permet de définir a^x lorsque l'exposant x est *négatif*, c'est-à-dire à définir l'exponentielle comme application $f : \mathbf{Z} \to \mathbf{Q}$.

On a en effet :

$$x + (-x) = 0$$

Donc :

$$f(x + (-x)) = f(0) = 1$$

Or :

$$f(x + (-x)) = f(x) \times f(-x)$$

Donc :

$$f(x) \times f(-x) = 1$$

$$\boxed{f(-x) = \frac{1}{f(x)}}$$

Soit, en notation a^x :

$$\boxed{a^{-x} = \frac{1}{a^x}}$$

d) *Exposants rationnels*

Peut-on donner un sens à a^x lorsque l'exposant x n'est plus entier (positif ou négatif) mais rationnel quelconque ? Autrement dit, peut-on définir l'exponentielle comme application dont le *départ* est **Q** ?

Quel sens faut-il par exemple, donner à $a^{1/2}$?

On a : $1/2 + 1/2 = 1$.

Donc :

$$f(1/2 + 1/2) = f(1) = a^1 = a$$

Or :

$$f(1/2 + 1/2) = f(1/2) \times f(1/2) = (f(1/2))^2$$

Donc $f(1/2)$ est la *racine carrée de a*.

$$\boxed{a^{1/2} = \sqrt{a}}$$

D'une façon générale, on démontre de même que :

$$f\left(\frac{1}{n}\right) = a^{1/n} = \sqrt[n]{a} \quad \text{(racine } n^{\text{e}} \text{ de } a\text{)}$$

Et :

$$f(p/q) = a^{p/q} = \sqrt[q]{a^p} = (\sqrt[q]{a})^p$$

Mais on a vu (chap. I, § 5) qu'en général $\sqrt{a}$, même lorsque a est *entier ou rationnel*, est un nombre *irrationnel*. Ceci nous amène à prendre comme ensemble d'*arrivée* de la fonction exponentielle l'ensemble des *nombres réels*. Comme les valeurs prises par les images $f(x) = a^x$ sont toujours positives, on prend comme ensemble d'arrivée l'ensemble $\mathbf{R}^*$ des *réels positifs*.

Ceci montre en quoi la justification de l'introduction des nombres réels se trouve dans la commodité que cela fournit pour les calculs *algébriques* : on ne peut faire de calcul exponentiel que dans la structure des réels, dès que l'exposant est rationnel, et non entier.

e) *Exponentielle en nombres réels*

On admettra que l'exposant peut lui aussi être non seulement entier ou rationnel, mais réel positif ou négatif quelconque, et que l'on peut définir l'*exponentielle de base* a ($a > 0$), comme une application :

$$f : \mathbf{R} \to \mathbf{R}^*$$

donnée par la formule :

$$\forall x \in \mathbf{R}, \quad f(x) = a^x \qquad (a \in \mathbf{R}^*)$$

Pour l'exponentielle ainsi définie, les *règles algébriques de calcul*, à connaître par cœur, sont celles que nous avons vues :

$a^0 = 1$	$f(0) = 1$
$a^1 = 1$	$f(1) = a$
$a^{x+y} = a^x \times a^y$	$f(x+y) = f(x) \times f(y)$
$a^{-x} = \frac{1}{a^x}$	$f(-x) = \frac{1}{f(x)}$

Des deux dernières, on déduit :

$$a^{x-y} = a^x \times a^{-y} = \frac{a^x}{a^y}$$

$$\boxed{a^{x-y} = \frac{a^x}{a^y}}$$

De ces règles, la plus fondamentale, parce qu'elle permet de retrouver les autres, est bien entendu celle qui traduit la propriété intrinsèque de la fonction exponentielle :

$$\boxed{a^{x+y} = a^x \times a^y}$$

Il convient d'y ajouter la règle :

$$\boxed{(a^x)^y = a^{x \times y}}$$

C'est la deuxième règle fondamentale du calcul exponentiel.

Démonstration dans le cas où x et y sont *entiers* naturels :

$$(a^x)^y = \underbrace{a^x \times a^x \times \ldots \times a^x}_{y \text{ fois}}$$

$$= \underbrace{\overbrace{a \times \ldots \times a}^{x \text{ fois}} \times \overbrace{a \ldots \times a}^{x \text{ fois}} \times \ldots \times \overbrace{a \times \ldots \times a}^{x \text{ fois}}}_{y \text{ fois}}$$

$$= \underbrace{a \times a \times \ldots \times a}_{(x \times y) \text{ fois}} = a^{x \times y}$$

Comme $x \times y = y \times x$, on en déduit :

$$(a^x)^y = (a^y)^x$$

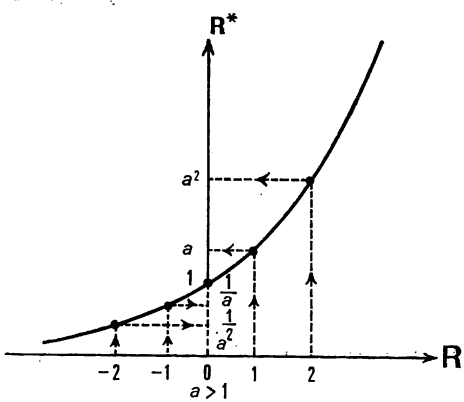
Attention aux parenthèses : $(a^x)^y$ n'*est pas* $a^{(x^y)}$

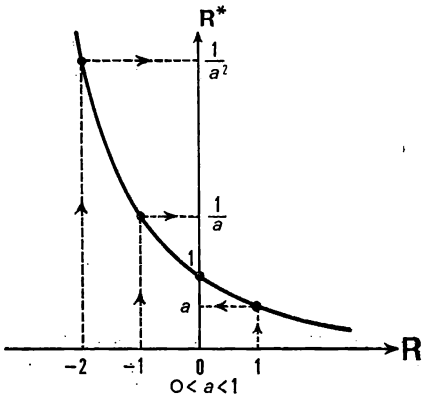
Exemple :

$$(2^3)^2 = 2^{3 \times 2} = 2^6 = 64$$

$$2^{(3^2)} = 2^9 = 512$$

Les *graphiques* sont :





$$f(x) = a^x$$

— Les deux *propriétés fondamentales*, équivalentes à la formule $f(x) = a^x$ $(a > 0)$, sont :

1° L'*image d'une somme* de deux nombres est *le produit des images :* l'addition est transformée en multiplication :

$$f(x + y) = f(x) \times f(y)$$
$$a^{x+y} = a^x \times a^y$$

2° L'ordre est conservé $(a > 1)$:

$$x > y \Rightarrow f(x) > f(y) \qquad (a^x > a^y)$$

ou

2° *bis.* L'ordre est inversé $(0 < a < 1)$:

$$x > y \Rightarrow f(x) < f(y) \qquad (a^x < a^y)$$

Comme on l'a vu au chap. I, § 5, $(\mathbf{R}, +, >)$ constitue un *groupe abélien ordonné* ; on a vu qu'il en est de même de $(\mathbf{R}^*, \times, >)$. Les fonctions exponentielles, qui appliquent $\mathbf{R}$ dans $\mathbf{R}^*$, en transformant $+$ en $\times$, et en conservant l'ordre, sont les *isomorphismes* de $(\mathbf{R}, +, >)$ dans $(\mathbf{R}^*, \times, >)$: elles transforment la première structure dans la seconde, et sont, on l'admet ici, *bijectives*.

$$(\mathbf{R}, +, >) \xrightarrow[\text{exp.}]{a^{\square}} (\mathbf{R}^*, \times, >)$$

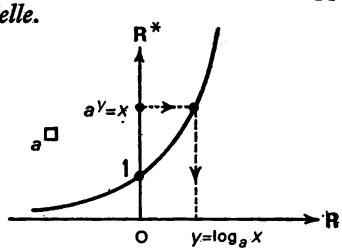
Ceci explique en particulier que 0, élément neutre de la structure de *départ* (addition), ait pour image 1, élément neutre de la structure d'*arrivée* (multiplication).

3. La fonction logarithme

a) *Fonction inverse de l'exponentielle*

La fonction exponentielle transforme des sommes en produits. Il serait avantageux d'avoir inversement une fonction qui transforme les produits en sommes : pour effectuer des multiplications, il suffirait de faire des additions.

Cette fonction va être définie comme *application inverse de l'exponentielle.*



Soit x un nombre *positif*, cherchons un nombre y tel que $a^y = x$ où a est une base donnée. On admettra qu'un tel nombre existe et est unique ; on le note $y = \log_a x$ qui se lit :

logarithme de base a de x. Ce nombre est l'*exposant* y de x écrit sous forme d'exponentielle de base a.

L'application $g : \mathbf{R}^* \to \mathbf{R}$ qui à chaque $x \in \mathbf{R}^*$ fait ainsi correspondre $\log_a x$ est appelée *fonction logarithme de base* a. C'est l'application *inverse* de l'*exponentielle de base* a.

Logarithme de base a, $\log_a$: application inverse de l'exponentielle de base a.

Donc par définition (le logarithme, c'est l'*exposant*) :

$$y = \log_a x \Leftrightarrow x = a^y \qquad (1)$$

Puisque la fonction $\log_a x$ est inverse de la fonction exponentielle, les propriétés de cette fonction s'établissent à partir de celles de la fonction exponentielle (1).

On a vu que : $a^0 = 1$.

Donc en appliquant la définition (1) avec : $x = 1, y = 0$:

$$\log_a 1 = 0$$

De même, on a vu que : $a^1 = a$.

Appliquons la définition (1) avec : $y = 1$ et $x = a$:

$$\log_a a = 1.$$

De même, il résulte de cette définition que :

$$\log_a a^x = x$$

(1) En particulier, comme $a^y > 0$ quel que soit y, le logarithme de x n'est défini *que si* x *est positif*.

Le logarithme de base a de a puissance x, c'est l'*exposant x*.

Le logarithme d'un produit est-il bien la somme des logarithmes ?

Soit z le nombre $\log_a (x \times y)$:

$$z = \log_a (x \times y) \Leftrightarrow x \times y = a^z$$

De même, posons :

$$u = \log_a x, \qquad v = \log_a y$$
$$u = \log_a x \Leftrightarrow x = a^u$$
$$v = \log_a y \Leftrightarrow y = a^v$$

Or d'après les propriétés fondamentales de la fonction exponentielle : $x \times y = a^u \times a^v = a^{u+v}$.

Mais :

$$x \times y = a^z$$

donc :

$$a^{u+v} = a^z$$

ce qui implique (puisque l'exponentielle est *bijective* entre **R** et **R***) :

$$u + v = z$$

Donc en remplaçant u, v, z par leurs valeurs :

$$\boxed{\log_a (x \times y) = \log_a x + \log_a y}$$

La fonction logarithme est donc bien la fonction cherchée. Elle transforme les produits en sommes.

On démontre de même que :

$$\boxed{\log_a x/y = \log_a x - \log_a y}$$

En particulier :

$$\log_a 1/x = -\log_a x$$

On démontre aussi que (prouvez le lorsque y est entier, puis lorsque y est rationnel) :

$$\log_a x^y = y \log_a x$$

b) *Propriétés caractéristiques de la fonction logarithme*

Comme l'exponentielle, la fonction logarithme est caractérisée par ses propriétés intrinsèques, qui résultent d'ailleurs de sa définition comme inverse de la fonction exponentielle. A savoir : la fonction logarithme est un *isomorphisme* du groupe abélien ordonné $(\mathbf{R}^*, \times, >)$ des réels positifs par rapport à la multiplication dans le groupe abélien ordonné $(\mathbf{R}, +, >)$ des réels positifs et négatifs par rapport à l'addition.

$$f : (\mathbf{R}^*, \times, >) \to (\mathbf{R}, +, >)$$

avec :

1° $f(x \times y) = f(x) + f(y)$

L'image d'un produit est la somme des images.

2° $x > y \Rightarrow f(x) > f(y)$ (base $a > 1$)

L'ordre est conservé

ou

2° *bis.* $x > y \Rightarrow f(x) < f(y)$

L'ordre est inversé (base $0 < a < 1$).

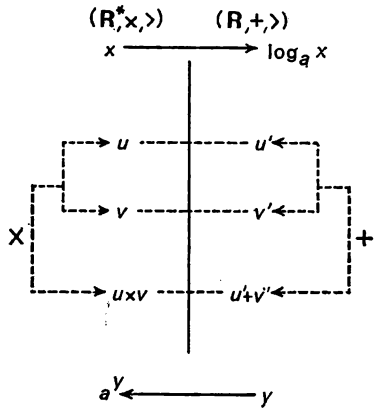
La base a du logarithme est le nombre a tel que $f(a) = 1$, c'est-à-dire dont l'image est 1.

Intuitivement, on peut regarder exponentielle et logarithme comme les applications qui permettent de passer du langage « additif » au langage « multiplicatif » ; tout ce qui se dit au moyen d'additions peut être traduit (par l'exponentielle) en termes de multiplications, et inversement (par le logarithme). Exponentielle et logarithme sont les deux faces du même *isomorphisme* entre $(\mathbf{R}, +, >)$ et $(\mathbf{R}^*, \times, >)$, selon qu'on le considère de $\mathbf{R}$ vers $\mathbf{R}^*$ ou de $\mathbf{R}^*$ vers $\mathbf{R}$.

Ce que résume le diagramme :

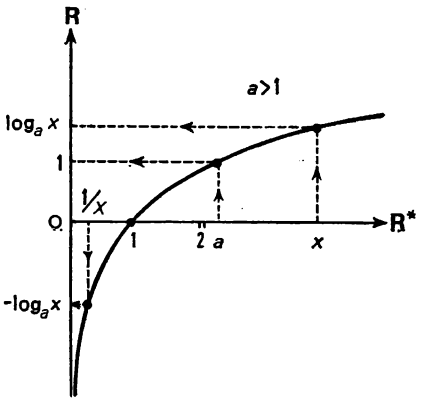
$$(\mathbf{R}, +, >) \underset{\text{log. } \log_a \square}{\overset{\text{exp. } a^{\square}}{\rightleftarrows}} (\mathbf{R}^*, \times, >)$$

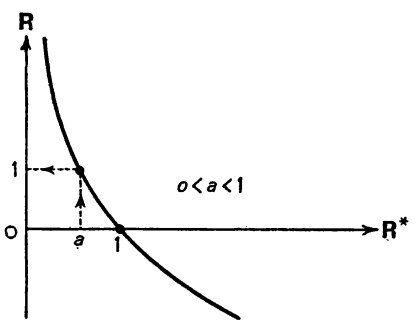
Ou celui-ci, conforme à la disposition des tables de logarithmes (cf. § 7) :



c) *Graphiques*

Le graphique de la fonction logarithme se déduit, d'après la définition, de celui de l'exponentielle en *échangeant le rôle des axes.*





4. La fonction puissance

Si a est un nombre entier naturel ($a \in \mathbf{N}$), on appelle fonction puissance d'*exposant* a la fonction définie par :

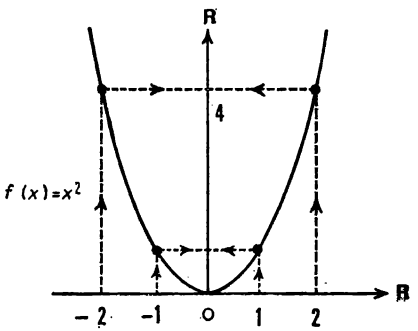
$$f(x) = x^a$$

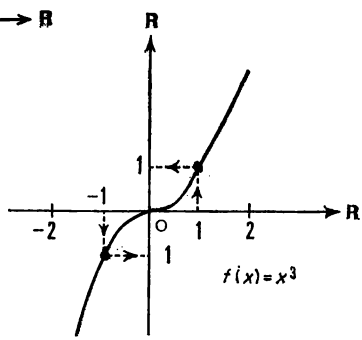
On peut la considérer comme une application f de $\mathbf{R}$ dans $\mathbf{R}$.

Exemples : $f(x) = x^2$ $(a = 2)$

$f(x) = x^3$ $(a = 3)$

dont les graphiques sont donnés ci-dessous.





— Si a est un nombre réel quelconque, on a vu (p. 59) qu'à condition que x soit un nombre *positif*, on peut définir x^a.

La formule :

$$\boxed{f(x) : x^a}$$

permet donc de définir une application :

$$f : \mathbf{R}^* \to \mathbf{R}^*$$

Exemple :

Pour $a = 1/2$, on a l'application définie par :

$$f(x) = x^{1/2} \; (= \sqrt{x})$$

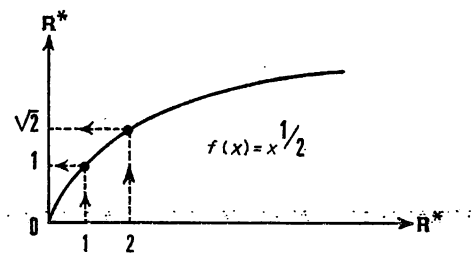
Comme :

$$(x^{1/2})^2 = x^{(1/2)\times 2} = x^1 = x$$

cette fonction est *l'inverse* de la fonction *carré* :

$$f(x) = x^2$$

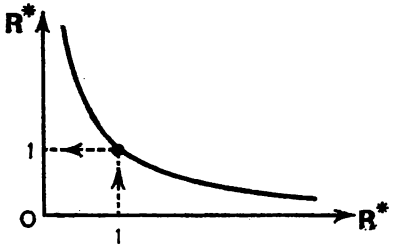
Son graphique se déduit de celui de x^2 en échangeant le rôle des axes (et en ne conservant que la partie *positive* de ceux-ci).



— L'exposant peut également être négatif.

Exemples :

$$a = -2, \qquad f(x) = x^{-2} = 1/x^2$$
$$a = -1, \qquad f(x) = x^{-1} = 1/x$$



— Les *propriétés caractéristiques* des fonctions puissance sont :

1° Que *l'image d'un produit est le produit des images* :

$$\boxed{(x \times y)^a = x^a \times y^a}$$

Cette propriété se démontre simplement si a est entier :

$$(x \times y)^a = \underbrace{(x \times y) \times (x \times y) \times \ldots \times (x \times y)}_{a \text{ fois}}$$
$$= \underbrace{(x \times \ldots \times x)}_{a \text{ fois}} \times \underbrace{(y \times \ldots \times y)}_{a \text{ fois}}$$
$$= x^a \times y^a$$

On l'admettra pour les autres valeurs de a (voir exercice 1, § *5).

2° Elle *conserve l'ordre* si $a > 0$:

$$x > y \Rightarrow x^a > y^a$$

(voir les graphiques $a = 2$, $a = 3$, $a = 1/2$).

Elle inverse l'ordre si $a < 0$:

$$x > y \Rightarrow x^a < y^a$$

(voir les graphiques pour $a = -2$, $a = -1$).

Ces propriétés sont *caractéristiques* en ce sens que toute application $f : \mathbf{R}^* \to \mathbf{R}^*$ telle que :

1°	$f(x \times y) = f(x \times y)$
2°	$x > y \Rightarrow f(x) > f(y)$
ou	
2° *bis.*	$x > y \Rightarrow f(x) < f(y)$

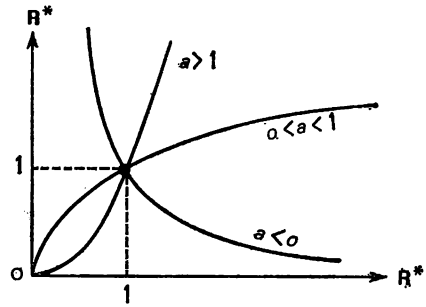
est une fonction puissance. Les fonctions puissance sont les *automorphismes* du groupe abélien ordonné $(\mathbf{R}^*, \times, >)$.

De même que la fonction *linéaire* $f(x) = a \times x$ a pour inverse la fonction *linéaire* $f^{-1}(x) = 1/a \times x$, la fonction puissance $f(x) = x^a$ a pour inverse, si $a \neq 0$, la fonction puissance $f^{-1}(x) = x^{1/a}$.

Ces propriétés sont résumées par le diagramme :

$$(\mathbf{R}^*, \times, >) \underset{\square^{1/a}}{\overset{\square^a}{\underset{\text{puiss.}}{\rightleftarrows}}} (\mathbf{R}^*, \times, >)$$

Les graphiques possibles pour ces fonctions sont inscrits page 66 sur une même figure.



On remarque que dans tous les cas : $f(1) = 1$.

En effet, 1, élément neutre de $\times$, doit être conservé par les applications qui conservent la structure multiplicative.

N. B. — L'égalité $(x \times y)^a = x^a \times y^a$ constitue la troisième des égalités à retenir par cœur, après les égalités $a^x \times a^y = a^{x+y}$ et $(a^x)^y = a^{x \times y}$ pour être capable de retrouver toutes les formules de calcul faisant intervenir des exposants et des logarithmes.

*5. Les quatre types fondamentaux de fonctions numériques

On a vu au chapitre I qu'il y a plusieurs espèces de nombres, plusieurs structures numériques ; toutes ont en commun d'être *totalement ordonnées*, et qu'on peut y faire des additions et des multiplications.

Dès lors, on comprend que parmi les applications dont départ et arrivée sont des ensembles de nombres, les plus fondamentales sont celles qui s'accordent bien avec les structures en jeu : le sens précis qu'il faut donner à cette expression est montré par l'exemple des 4 fonctions numé-

riques fondamentales sur **R**, la plus riche des structures numériques en possibilités de calcul algébrique.

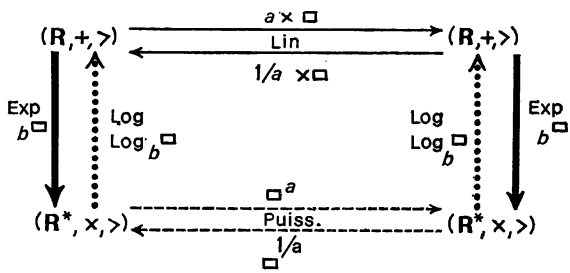
C'est, d'une part, de conserver l'ordre (ou de l'inverser), d'autre part, soit de conserver l'une des opérations :

$$-(+ \to +), (\times \to \times)$$

soit de transformer l'une en l'autre :

$$-(+ \to \times), (\times \to +)$$

Le diagramme ci-dessous résume les propriétés des 4 fonctions par rapport aux structures numériques.



Les autres fonctions numériques usuelles sont des combinaisons des 4 fonctions fondamentales :

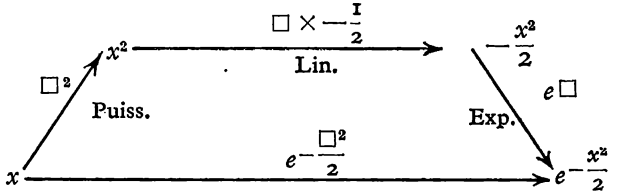
Linéaire	$+ \to +$
Exponentielle.....	$+ \to \times$
Logarithme	$\times \to +$
Puissance	$\times \to \times$

C'est le cas par exemple de :

— la fonction affine — $a \times x + b$ — somme d'une fonction constante et d'une fonction linéaire ;

— la fonction $f(x) = ax^2 + bx + c$, combinaison linéaire de fonctions puissance ;

— la fonction $f(x) = e^{-x^2/2}$, qui intervient en calcul des probabilités et en statistique, qui est décomposable selon le diagramme :



Un seul type de fonctions usuelles semble faire exception : les fonctions périodiques, comme par exemple $f(x) = \sin x$; en fait, il n'en est rien. Dans la structure numérique convenable, et que l'on n'étudie pas ici, celle des « nombres complexes », les fonctions trigonométriques sont des combinaisons linéaires de fonctions exponentielles.

Exercices :

1. Montrer que si $y = x^a$, alors $\log_b y = a \times \log_b x$, et réciproquement que si $y = a \times x$, alors $b^y = (b^x)^a$. En déduire que les fonctions puissance sont les seuls automorphismes de $(\mathbf{R}^*, \times, >)$.

2. Prouver que la seule application $f : \mathbf{Q} \to \mathbf{Q}$ telle que $f(x + y) = f(x) + f(y)$ et $f(x \times y) = f(x) \times f(y)$, quels que soient x et y, est l'application identique :

$$\forall x \in \mathbf{Q}, \qquad f(x) = x$$

Quel est l'ensemble des automorphismes du corps $(\mathbf{Q}, +, \times)$?

Remarque :

L'arithmétique élémentaire traite usuellement des « quatre opérations » définies sur le corps $\mathbf{Q}$ des rationnels : $+$ et son inverse $-$; $\times$ et son inverse :.

On a vu que la recherche des morphismes entre structure additive et structure multiplicative conduisait à introduire :

1° une structure plus riche, complétant $\mathbf{Q} : \mathbf{R}$;

2° deux nouvelles opérations dans $\mathbf{R}$: l'exponentielle (x^y) et son inverse le logarithme $(\log_x y)$, toujours définies si x et y sont positifs.

Pour $\mathbf{R}$ muni de ses six opérations, les morphismes de $+$ à exp., ou de exp. à exp., etc., n'introduisent pas de nouvelles opérations : ces morphismes se réduisent à des fonctions constantes, ou à l'application identique.

Dans le même ordre d'idées, le seul automorphisme de la structure d'anneau de $(\mathbf{Z}, +, \times)$, ou de $(\mathbf{D}, +, \times)$ comme de celle de corps de $\mathbf{Q}$ $(+, \times)$ et de $(\mathbf{R}, +, \times)$ est l'application identique (cf. exercice *supra*).

6. Application : les papiers fonctionnels

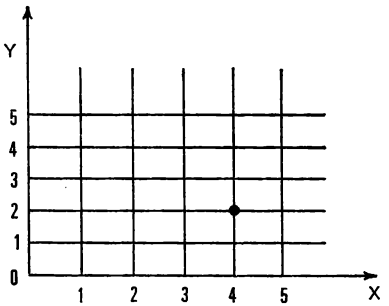
Une application immédiate de ce qu'on vient d'apprendre sur les 4 types de fonctions numériques est constituée par les « papiers fonctionnels ».

a) *Ajustement sur une droite : papier à échelles « arithmétiques »*

On veut savoir si la liaison entre deux « variables numériques » X et Y est, au moins approximativement, de la forme *affine* :

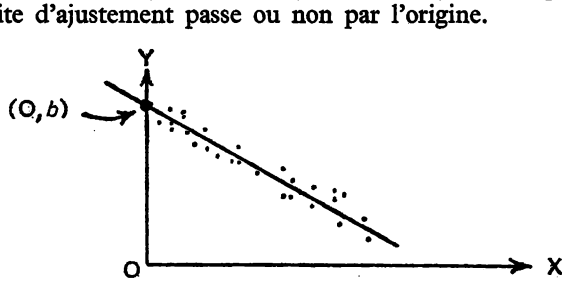
$$Y = aX + b$$

On se munit de papier millimétré dit « en échelles arithmétiques » : les points de division équidistants sur chaque axe sont gradués selon une progression arithmétique, par exemple 0, 1, 2, etc.



Puis, on marque sur le papier tous les couples de valeurs observées ; par exemple, si l'on a fait une observation dans laquelle x vaut 4 et y vaut 2, le couple est représenté par le point marqué sur la figure ci-dessus ; c'est le couple (4, 2).

Si le « nuage de points » obtenu est tel que ces points soient grossièrement alignés, la liaison est bien approximativement linéaire ($y = ax$), ou affine ($y = ax + b$), selon que la droite d'ajustement passe ou non par l'origine.



On verra ultérieurement comment ajuster « au mieux » une droite sur un nuage de points (chap. V). Graphiquement, on s'efforce de faire en sorte qu'il y ait à peu près autant de points de part et d'autre de la droite.

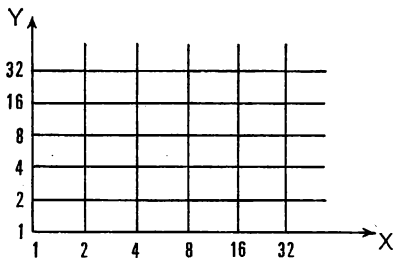
La droite une fois tracée, sa pente donne la valeur de a, et son « ordonnée à l'origine » celle de b (voir la figure).

b) *Ajustement sur une fonction « puissance » : papier à échelles bilogarithmiques*

Si la liaison que l'on soupçonne entre X et Y est une fonction puissance $Y = X^a$, ou plus généralement $Y = bX^a$, on utilise le fait que dans ce cas, en prenant les logarithmes dans une base arbitraire : $\log Y = \log b + a \log X$. C'est-à-dire que la liaison entre log X et log Y doit être de la forme affine.

Donc les couples de valeurs observées doivent être approximativement alignés si on les représente sur du papier millimétré gradué non comme les valeurs prises, mais comme leurs logarithmes : le papier à échelles logarithmiques (ou papier bilogarithmique).

Il est facile de construire un tel papier ; il suffit de prendre du papier millimétré ordinaire, mais de graduer les points de division *équidistants* de chaque axe selon les termes d'une progression géométrique (selon une fonction exponentielle).



Sur la figure, on a utilisé l'exponentielle ($f(x) = 2^x$) de base 2 pour indiquer les graduations.

Le commerce vend des papiers dans lesquels des graduations intermédiaires sont également indiquées, correspondant à des valeurs allant en progression arithmétique : d'où l'effet de resserrement.

Si, sur ce papier, les points sont approximativement ali-

gnés, on peut estimer qu'on est en présence d'une liaison de type :

$$Y = bX^a$$

L'exposant a est d'ailleurs donné par la pente de la droite d'ajustement :

$$\log Y = \log b + a \log X$$

Le coefficient b est calculé à partir de son logarithme : $\log b$ est l'ordonnée à l'origine de la droite d'ajustement.

c) *Ajustement sur une exponentielle (un logarithme) : papier à échelles « semi-logarithmiques »*

Si enfin on soupçonne une liaison du type exponentiel :

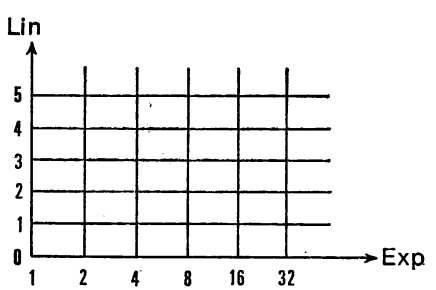
$$Y = b.a^x \qquad (\log Y = \log b + X \log a)$$

ou logarithmique :

$$Y = A \log X + B \qquad (X = \beta.\alpha^Y)$$

on utilise du papier à échelles « semi-logarithmiques ».

L'un des axes est gradué selon une progression arithmétique (fonction linéaire), l'autre selon une progression géométrique (fonction exponentielle). Si les points sont alignés dans ce système d'axes, celle des « variables » correspondant à l'axe gradué en progression géométrique est fonction exponentielle de celle qui correspond à l'axe gradué arithmétiquement.



De façon précise : si la liaison est $Y = b.a^x$, les valeurs de Y correspondent, dans la figure ci-dessus, à l'axe horizontal, celles de X à l'axe vertical ; si la liaison est $Y = A \log X + B$, ce sont les valeurs de X qui sont portées sur l'axe horizontal, celles de Y sur l'axe vertical.

N.B. — L'avantage dans l'utilisation de ces papiers consiste en ce que le problème d'ajustement se ramène toujours à celui d'une droite par rapport à un nuage de points. Ce dernier problème est simple ; il est facile de juger « à vue » si des points sont alignés (approximativement) ou non ; l'instrument qu'est la règle y aide. Alors qu'il est difficile de voir si des points s'ajustent sur une fonction exponentielle ou une fonction puissance (comparer les graphiques de ces fonctions ; sur certaines de leurs parties, les courbes se ressemblent beaucoup).

En ce qui concerne les méthodes algébriques d'ajustement, c'est également l'ajustement sur une droite (fonction affine) qui fournit des calculs commodes (chap. V).

7. Calcul pratique de logarithmes

Étant donné différents nombres $x_1, x_2, \ldots$, on a souvent besoin de calculer $f(x_1), f(x_2) \ldots$ Si la fonction donnée est linéaire ($f(x) = ax$) ou une fonction puissance simple ($f(x) = x^2$ par exemple) le calcul est aisé ; le seul outil à employer est la multiplication. Mais si l'on veut calculer au moins une valeur approchée de $2^{3,2}$ ou de $(4{,}653)^{3,2}$, il faut avoir recours à un autre outil, par exemple la table de logarithmes.

a) *De l'usage des tables de logarithmes*

— Les tables de logarithmes qui se vendent dans le commerce donnent, pour les nombres de 1 000 à 10 000,

la partie *décimale* de $\log_{10} x$, appelé logarithme décimal et que l'on notera simplement $\log x$.

— Pourquoi ne donner que la partie décimale (c'est-à-dire celle qui est après la virgule) d'un nombre ? C'est que la partie avant la virgule, dite partie entière, peut être connue très facilement ; en effet, prenons un exemple ; on a :

$$\log 2 = 0{,}30103$$

Cherchons $\log 20$, $\log 2\,000$, $\log 0{,}02$:

$$\begin{aligned} 20 &= 2 \times 10 \\ 2\,000 &= 2 \times 10^3 \\ 0{,}02 &= 2 \times 10^{-2} \end{aligned}$$

Donc, d'après les propriétés des logarithmes (§ 3) :

$$\begin{aligned} \log 20 &= \log 2 + \log 10 \\ \log 2\,000 &= \log 2 + \log 10^3 = \log 2 + 3 \log 10 \\ \log 0{,}02 &= \log 2 + \log 10^{-2} = \log 2 - 2 \log 10 \end{aligned}$$

Mais comme il s'agit du logarithme de base 10 :

$$\log 10 = 1$$

D'où

$$\begin{aligned} \log 20 &= 0{,}30103 + 1 = 1{,}30103 \\ \log 2\,000 &= 0{,}30103 + 3 = 3{,}30103 \\ \log 0{,}02 &= 0{,}30103 - 2 = ? \end{aligned}$$

Pour éviter que la partie après la virgule soit négative, on n'effectue pas cette soustraction, car, dans la suite des calculs, cela serait incommode ; aussi on note $\log 0{,}02 = \bar{2}{,}30103$ ce qui veut dire que le signe — n'affecte que la partie entière, la partie après la virgule restant positive : $\bar{2}{,}30103 = -2 + 0{,}30103$. Grâce à cet artifice et aux formules qui donnent les propriétés des logarithmes, on peut effectuer tous les calculs.

b) *Un exemple de calcul*

Calculer le nombre : $A = \dfrac{(9\,537)^{0,4} \times (0,06)^{1,5}}{48,3}$.

On peut écrire :

$$\log A = 0,4 \log 9\,537 + 1,5 \log 0,06 - \log 48,3$$

La partie entière de $\log 9\,537$ est 3.
— $\log 0,06$ - — 2.
— $\log 48,3$ - 1.

On lit dans la table la partie décimale de chacun de ces logarithmes :

La partie décimale de $\log 9\,537$ est 97 941.
— $\log 0,06$ - 77 815.
— $\log 48,3$ - 68 395.

On peut donc écrire :

$$\begin{aligned} \log 9\,537 &= 3,97941 \\ \log 0,06 &= \bar{2},77815 \\ \log 48,5 &= 1,68395 \end{aligned}$$

Calculons :

$$\begin{aligned} 0,4 \log 9\,537 &= 0,4 \times 3,97941 \\ &= 1,591764 \\ 1,5 \log 0,06 &= 1,5 \times (-2 + 0,77815) \\ &= -3 + 1,5 \times 0,77815 \\ &= -3 + 1,16722 \\ &= -2 + 0,16722 \\ &= \bar{2},16722 \\ -\log 48,3 &= -1,68395 \\ &= -1 - 0,68395 \\ &= -2 + 1 - 0,68395 \\ &= -2 + 0,31605 \\ &= \bar{2},31605 \end{aligned}$$

Et, pour connaître log A, on fait l'addition prescrite :

$$\begin{aligned} 0{,}4 \log 9\,537 &= 1{,}59176 \\ 1{,}5 \log 0{,}06 &= \bar{2}{,}16722 \\ -\log 48{,}3 &= \bar{2}{,}31605 \\ \hline \log A &= \bar{2}{,}07503 \end{aligned}$$

On a additionné les parties décimales entre elles puisqu'elles sont toutes positives ; pour la partie entière du résultat, il vient 1 de retenue de la partie décimale, plus $(1 + (-2) + (-2))$; le résultat est donc -2.

Puisque la partie entière de log A est -2 :

$$A = x \times 10^{-2}$$

où x est un nombre tel que $1 < x \leqslant 10$.

Pour trouver x, on retourne à la table de logarithme, mais en la lisant à l'envers ; c'est-à-dire que l'on va chercher le nombre x dont la *partie décimale* de son logarithme se rapproche le plus de 07503.

On lit :

$$\begin{aligned} \log 1{,}188 &= 0{,}07482 \\ \log 1{,}189 &= 0{,}07518 \end{aligned}$$

Le second résultat est le plus près de celui que l'on cherche, donc, *approximativement* :

$$x \simeq 1{,}189 \quad \text{(par excès)}$$

et on peut donc écrire, comme valeur approximative de A :

$$\boxed{A \simeq 0{,}01189}$$

N.B. — Toute table de logarithmes est en même temps une table d'exponentielles ; la lecture dans le sens « à l'envers » de la table de logarithme de base 10, comme on vient de le faire ci-dessus, est une lecture de l'exponentielle de base 10 (cf. second diagramme du § 3, *b*).

Récapitulation

Quatre fonctions numériques fondamentales :
Toutes conservent, ou inversent l'*ordre* des nombres :

$$> \to > \text{ ou } > \to <$$

— Linéaire	$(\mathbf{R}, +) \to (\mathbf{R}, +)$	$a \times x$
— Exponentielle	$(\mathbf{R}, +) \to (\mathbf{R}^*, \times)$	a^x
— Logarithme	$(\mathbf{R}^*, \times) \to (\mathbf{R}, +)$	$\log_a x$
— Puissance	$(\mathbf{R}^*, \times) \to (\mathbf{R}^*, \times)$	x^a

Les trois formules à partir desquelles peuvent se faire tous les calculs exponentiels :

$$\begin{aligned} a^{b+c} &= a^b \times a^c \\ (a^b)^c &= a^{b \times c} \\ a^b \times c^b &= (a \times c)^b \end{aligned}$$

Passage de l'exponentielle au logarithme (le *logarithme*, c'est l'*exposant*) :

$$c = a^b \Leftrightarrow b = \log_a c$$

DEUXIÈME PARTIE

LA MESURE EN DESCRIPTION STATISTIQUE

CHAPITRE III

PRODUIT SCALAIRE

Dans cette partie, on aborde la mesure en description statistique (statistique, c'est-à-dire par dénombrement) ; on sait que le cadre algébrique dans lequel se situent la plupart des concepts et des méthodes de la description statistique est celui de l'algèbre linéaire *; la notion de base est celle de* vectoriel. *Pour l'introduction à cette notion on renvoie à* MSH, *III*e *partie, qui lui est consacrée ; c'est là que l'étudiant retrouvera les définitions de : vecteur, scalaire, moyenne, forme linéaire, base d'un vectoriel, toutes indispensables à l'intelligence de ce qui suit, dans le mode d'exposition adopté ici.*

On rappelle une fois pour toutes que les vecteurs dont on parle ici peuvent toujours être interprétés comme des suites de nombres ou statistiques *; abstraitement, ce sont des applications d'un ensemble fini* $\Omega = \{1, 2, \ldots, n\}$ *dans un* corps *de nombres (les scalaires), qui sera toujours* $\mathbf{R}$ *ou* $\mathbf{Q}$ *(chap. I, § 4 et 5). Une telle application* $\mathbf{x} : \Omega \to \mathbf{R}$ *(ou* $\mathbf{Q}$*) est notée :* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ *; pour tout* $i \in \Omega$, $\mathbf{x}_i$ *désigne son image, le nombre* $\mathbf{x}_i \in \mathbf{R}$ *qui lui correspond dans l'application* $\mathbf{x}$.

L'ensemble $\mathbf{R}^\Omega$ *(ou* $\mathbf{R}^n$*) de toutes les applications de* Ω *dans* $\mathbf{R}$ *est muni d'une structure de vectoriel en définissant l'*addition *de deux vecteurs (applications)* $\mathbf{x}$ *et* $\mathbf{y}$ *par :*

$$\mathbf{x} + \mathbf{y} = (\mathbf{x}_1 + \mathbf{y}_1, \ldots, \mathbf{x}_n + \mathbf{y}_n)$$

*et l'*homothétie *dans le rapport* $\lambda \in \mathbf{R}$ *(le produit par* λ*) par :*

$$\lambda\mathbf{x} = (\lambda\mathbf{x}_1, \ldots, \lambda\mathbf{x}_n)$$

Les concepts que l'on va introduire dans cette partie sont ceux qui fournissent des mesures *de la* dispersion *d'une statistique (d'une suite de nombres ; d'une application dans* **R** *; d'un vecteur), de l'écart, de la* distance *entre deux statistiques, et de leur* corrélation.

L'outil essentiel de cette mesure, dans une structure de vectoriel, est le produit scalaire, *étudié dans le présent chapitre, et dont dérivent les autres outils : distance, variance, cosinus, coefficient de corrélation linéaire.*

Note sur la typographie :

Dans la suite, les caractères italiques désignent des nombres ; les caractères gras désignent des vecteurs.

Pour les opérations, on note + gras l'addition des *vecteurs*, et + maigre celle des nombres ; la multiplication des nombres est notée par ×, ou sans aucun signe, lorsqu'il n'y a pas risque de confusion ; le signe . désigne toujours le *produit scalaire* de *vecteurs*, jamais un produit de nombres.

1. Produit scalaire

a) *Définition*

Soient deux vecteurs :

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \text{ et } \mathbf{y} = (y_1, y_2, \ldots, y_n)$$

d'un *même vectoriel* de dimension n ; à ce couple de deux vecteurs, on fait correspondre le *nombre*, noté $\mathbf{x}.\mathbf{y}$ calculé par la règle :

$$\mathbf{x}.\mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

Le nombre $\mathbf{x}.\mathbf{y}$ est la somme des produits de chaque coordonnée x_i de $\mathbf{x}$ par la coordonnée correspondante y_i de $\mathbf{y}$.

Exemple :

Dans $\mathbf{R}^5$:

$$\mathbf{x} = (1, 0, -1, 2, 2) \qquad \mathbf{y} = (-2, 1, 3, -1, 2)$$
$$\mathbf{x}.\mathbf{y} = 1 \times (-2) + 0 \times 1 + (-1) \times 3 + 2 \times (-1) + 2 \times 2$$
$$= -2 + 0 - 3 - 2 + 4 = -3$$

L'*application* qui à chaque couple $(\mathbf{x}, \mathbf{y})$ de vecteurs d'un vectoriel V fait correspondre ainsi le nombre (un nombre et un seul) $\mathbf{x}.\mathbf{y}\ (= x_1 y_1 + \ldots + x_n y_n)$ est une application f du produit cartésien $\mathrm{V} \times \mathrm{V}$ (ensemble des couples de vecteurs de V) dans le corps des nombres, des *scalaires* de V ($\mathbf{R}$ ou $\mathbf{Q}$, selon les cas).

$$f : \mathrm{V} \times \mathrm{V} \to \mathbf{R}$$

Dans cette application, l'image d'un couple $(\mathbf{x}, \mathbf{y})$ de vecteurs de V est le scalaire (le nombre) :

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{x}.\mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

f est appelé *produit scalaire* sur V.

Il s'agit bien en effet d'un *produit* entre vecteurs, on fait le produit $\mathbf{x}.\mathbf{y}$ de *deux vecteurs* ; mais le résultat de ce produit est un *scalaire* (un nombre), et non un vecteur.

Produit scalaire sur le vectoriel $\mathbf{R}^n$ (resp. $\mathbf{Q}^n$) de dimension n sur le corps $\mathbf{R}$ des réels (resp. sur le corps $\mathbf{Q}$ des rationnels) : application $f : \mathrm{V} \times \mathrm{V} \to \mathbf{R}$ (resp. $\mathbf{Q}$) avec :

$$f(\mathbf{x}, \mathbf{y}) = \mathbf{x}.\mathbf{y} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

b) *Exemples et cas particuliers*

a) Soit dans $\mathbf{R}^3$ les deux vecteurs :

$$\mathbf{x} = (1, -1, 1)$$
$$\mathbf{y} = (2, 4, 2)$$
$$\mathbf{x}.\mathbf{y} = 2 - 4 + 2 = 0$$

Cet exemple montre que le *produit scalaire de deux vecteurs peut prendre la valeur o sans qu'aucun des deux vecteurs soit nul.*

b) Si $\mathbf{x} = (0, 0, 0) = \mathbf{0}$ et $\mathbf{y} = (y_1, y_2, y_3)$:

$$\mathbf{x}.\mathbf{y} = \mathbf{0}.\mathbf{y} = 0 \times y_1 + 0 \times y_2 + 0 \times y_3 = 0 + 0 + 0 = 0$$

> Si l'un des vecteurs est nul, le produit scalaire prend la valeur o.

c) Considérons deux des vecteurs de la *base* (de $\mathbf{R}^3$) constituée par les trois vecteurs :

$$\mathbf{e}_1 = (1, 0, 0) \qquad \mathbf{e}_2 = (0, 1, 0) \qquad \mathbf{e}_3 = (0, 0, 1)$$

On a :

$$\mathbf{e}_1.\mathbf{e}_2 = 1 \times 0 + 0 \times 1 + 0 \times 0 = 0$$

De même :

$$\mathbf{e}_2.\mathbf{e}_3 = \mathbf{e}_1.\mathbf{e}_3 = 0$$

Le produit scalaire de deux quelconques (et distincts) des vecteurs de cette base est nul ; pourtant, aucun de ces vecteurs n'est nul.

c) *Propriétés du produit scalaire : règles de calcul*

Pour fixer les idées, plaçons-nous dans le vectoriel $\mathbf{R}^3$ de dimension 3. Si $\mathbf{x}$ et $\mathbf{y}$ sont deux vecteurs, on a ainsi par définition :

$$\mathbf{x}.\mathbf{y} = x_1 y + x_2 y_2 + x_3 y_3$$

a) *Commutativité :*

Il est clair que le produit scalaire est *commutatif* ; changer l'ordre entre $\mathbf{x}$ et $\mathbf{y}$ ne change pas la valeur prise par le produit scalaire :

$$\mathbf{y}.\mathbf{x} = y_1 x_1 + y_2 x_2 + y_3 x_3 = x_1 y_1 + x_2 y_2 + x_3 y_3$$

$$\boxed{\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x}}$$

b) *Produit par un nombre. Homothéties :*

Multiplions le vecteur $\mathbf{x}$ par un scalaire λ :

Comme $\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \lambda x_3)$:

$$\begin{aligned}(\lambda\mathbf{x}.\mathbf{y}) &= (\lambda x_1) y_1 + (\lambda x_2) y_2 + (\lambda x_3) y_3 \\ &= \lambda (x_1 y_1 + x_2 y_2 + x_3 y_3) = \lambda (\mathbf{x}.\mathbf{y})\end{aligned}$$

$$\boxed{(\lambda\mathbf{x}).\mathbf{y} = \lambda (\mathbf{x}.\mathbf{y})}$$

De même :

$$\boxed{(\mathbf{x}.\mu\mathbf{y}) = \mu (\mathbf{x}.\mathbf{y})}$$

Multiplier l'un des vecteurs par un nombre multiplie par ce nombre la valeur prise par le produit scalaire.

Exemple :

$$\begin{aligned}(2\,\mathbf{x}).(6\,\mathbf{y}) &= 2\,(\mathbf{x}.6\,\mathbf{y}) = (2 \times 6)\,(\mathbf{x}.\mathbf{y}) \\ &= 12\,(\mathbf{x}.\mathbf{y})\end{aligned}$$

En particulier, si chacun des deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ est multiplié par un *même nombre* λ (subit une homothétie de rapport λ), le produit scalaire est multiplié par λ^2.

$$\boxed{(\lambda\mathbf{x}).(\lambda\mathbf{y}) = \lambda^2 (\mathbf{x}.\mathbf{y})}$$

c) *Distributivité :*

Soit $\mathbf{z}$ un troisième vecteur ; $\mathbf{z} = (z_1, z_2, z_3)$.
Produit scalaire de $\mathbf{x}$ avec le vecteur $(\mathbf{y} + \mathbf{z})$:

$$\mathbf{x}.(\mathbf{y} + \mathbf{z}) = x_1(y_1 + z_1) + x_2(y_2 + z_2) + x_3(y_3 + z_3)$$

puisque :

$$\mathbf{y} + \mathbf{z} = (y_1 + z_1, y_2 + z_2, y_3 + z_3)$$

Or :

$$\begin{aligned} x_1(y_1 + z_1) + x_2(y_2 + z_2) + x_3(y_3 + z_3) &= x_1 y_1 + x_1 z_1 + \ldots + x_3 y_3 + x_3 z_3 \\ &= (x_1 y_1 + x_2 y_2 + x_3 y_3) + (x_1 z_1 + x_2 z_2 + x_3 z_3) \\ &= \mathbf{x}.\mathbf{y} + \mathbf{x}.\mathbf{z} \end{aligned}$$

$$\boxed{\mathbf{x}.(\mathbf{y} + \mathbf{z}) = \mathbf{x}.\mathbf{y} + \mathbf{x}.\mathbf{z}}$$

Le produit scalaire d'une *somme* de vecteurs par un vecteur $\mathbf{x}$ est la somme des produits scalaires de chacun de ces vecteurs par $\mathbf{x}$. Le *produit scalaire est distributif par rapport à l'addition des vecteurs.*

On calcule sur le produit scalaire des vecteurs comme sur la multiplication des nombres (mais on n'a pas le droit de faire des divisions ; voir l'exercice d'application ci-dessous).

Exemples :

$$\mathbf{x}.(2\,\mathbf{y} + 3\,\mathbf{z}) = 2\,(\mathbf{x}.\mathbf{y}) + 3\,(\mathbf{x}.\mathbf{z})$$

$$(2\,\mathbf{x} - \mathbf{y}).(\mathbf{z} + 3\,\mathbf{t}) = 2\,(\mathbf{x}.\mathbf{z}) - (\mathbf{y}.\mathbf{z}) + 6\,(\mathbf{x}.\mathbf{t}) - 3\,(\mathbf{y}.\mathbf{t})$$

Exercice d'application :

$\mathbf{a}$ étant un vecteur donné et non nul, peut-on conclure de l'égalité :

$$\mathbf{a}.\mathbf{x} = \mathbf{a}.\mathbf{y}$$

que : $\mathbf{x} = \mathbf{y}$? Autrement dit, peut-on faire des simplifications (ici par $\mathbf{a}$) ?

Réponse :

Non ; $\mathbf{a}.\mathbf{x} = \mathbf{a}.\mathbf{y}$ équivaut à $\mathbf{a}.\mathbf{x} - \mathbf{a}.\mathbf{y} = 0$.
C'est-à-dire : $\mathbf{a}.(\mathbf{x} - \mathbf{y}) = 0$, d'après la distributivité.

Or on a vu des exemples dans lesquels le produit scalaire de deux vecteurs est nul sans qu'aucun de ces vecteurs soit nul ; donc on ne peut pas conclure $\mathbf{x} - \mathbf{y} = \mathbf{0}$, c'est-à-dire $\mathbf{x} = \mathbf{y}$.

Exemple :

$\mathbf{a} = (1, 2, 1)$. Vérifiez vous-même, en utilisant la définition du produit scalaire, que si :
$\mathbf{x} = (3, 1, 4)$ et $\mathbf{y} = (4, 0, 5)$, on a : $\mathbf{a}.\mathbf{x} = \mathbf{a}.\mathbf{y}$ $(= 9)$

d) *Positivité*

Que vaut le produit scalaire d'un vecteur avec lui-même ? Dans $\mathbf{R}^n$, on a :

$$\mathbf{x}.\mathbf{x} = x_1.x_1 + x_2.x_2 + \ldots + x_n.x_n$$
$$= x_1^2 + x_2^2 + \ldots + x_n^2$$

$\mathbf{x}.\mathbf{x}$ est donc la somme des carrés des coordonnées de $\mathbf{x}$. Comme des carrés sont toujours *positifs ou nuls*, on a :

$$\forall \mathbf{x} \in \mathbf{R}^n \qquad \boxed{\mathbf{x}.\mathbf{x} \geqslant 0}$$

Le produit scalaire d'un vecteur avec lui-même est toujours positif ou nul.

En particulier, $\mathbf{x}.\mathbf{x}$ n'est nul que si chacun des carrés est nul, c'est-à-dire si *toutes les coordonnées de* $\mathbf{x}$ sont nulles : donc si $\mathbf{x}$ est le vecteur nul $\mathbf{0} = (0, 0, \ldots, 0)$.

$$\boxed{\mathbf{x}.\mathbf{x} = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}}$$

*2. Caractérisation algébrique du produit scalaire

a) *Produit scalaire et formes linéaires*

Si l'on choisit dans le vectoriel $V = \mathbf{R}^n$ (ou $\mathbf{Q}^n$) un vecteur *fixé* $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, le produit scalaire de celui-ci avec un vecteur variable $\mathbf{x}$ de V est :

$$\mathbf{u}.\mathbf{x} = u_1 x_1 + u_2 x_2 + \ldots + u_n x_n$$

C'est une *forme linéaire* sur V (cf. *MSH*, chap. XIII), c'est-à-dire une application $u : V \to \mathbf{R}$ qui à chaque vecteur $\mathbf{x}$ fait correspondre un nombre (un scalaire) et un seul, le nombre $u(\mathbf{x}) = \mathbf{u}.\mathbf{x}$; ce nombre se calcule comme somme des coordonnées de $\mathbf{x}$ pondérée par les coefficients u_i, c'est-à-dire les coordonnées de $\mathbf{u}$.

Les propriétés vues *supra* en 1-*b*, 1-*c*, expriment d'ailleurs que :

1) $u(\lambda\mathbf{x}) = \mathbf{u}.(\lambda\mathbf{x}) = \lambda(\mathbf{u}.\mathbf{x}) = \lambda u(x)$;
2) $u(\mathbf{x}+\mathbf{y}) = \mathbf{u}.(\mathbf{x}+\mathbf{y}) = \mathbf{u}.\mathbf{x} + \mathbf{u}.\mathbf{y} = u(\mathbf{x}) + u(\mathbf{y})$,

propriétés qui, on le sait, caractérisent les formes linéaires.

Ainsi, étant donné le produit scalaire $\mathbf{x}.\mathbf{y}$ de *deux* vecteurs, fixer l'*un de ces deux vecteurs* définit une forme linéaire, une application *linéaire* du vectoriel dans son corps de scalaires.

On traduit cette propriété du produit scalaire en disant qu'il est *bilinéaire*.

b) *Formes bilinéaires*

Une forme bilinéaire sur un vectoriel V est une application $f : V \times V \to \mathbf{R}$ qui à chaque couple $(\mathbf{x}, \mathbf{y})$ de vecteurs associe un nombre et un seul $f(\mathbf{x}, \mathbf{y})$, de telle sorte que si l'un des deux vecteurs est fixé, f, qui n'est plus alors que fonction du second vecteur (celui qui n'est pas fixé), soit alors une forme linéaire.

Comme on vient de le voir, le produit scalaire $\mathbf{x}.\mathbf{y}$ de deux vecteurs est une forme bilinéaire.

En outre, comme $\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x}$, le produit scalaire est une forme bilinéaire *symétrique* c'est-à-dire satisfaisant à :

$$\forall \mathbf{x}, \forall \mathbf{y}, \qquad f(\mathbf{x}, \mathbf{y}) = f(\mathbf{y}, \mathbf{x})$$

Enfin, comme $\mathbf{x}.\mathbf{x} \geqslant 0$ (positivité) et comme $\mathbf{x}.\mathbf{x}$ n'est nul que si $\mathbf{x}$ est nul, le produit scalaire est une forme bilinéaire *positive*, c'est-à-dire satisfaisant à :

$$\forall \mathbf{x}, \quad f(\mathbf{x}, \mathbf{x}) \geqslant 0 \qquad \text{et} \qquad f(\mathbf{x}, \mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$$

On montre que, *réciproquement*, toute forme *bilinéaire*, *symétrique* et *positive* sur un vectoriel V de dimension n, peut, par rapport à *une base convenable* (cf. *MSH*, chap. XII), se calculer par la formule :

$$f(\mathbf{x}, \mathbf{y}) = p_1 x_1 y_1 + p_2 x_2 y_2 + \ldots + p_n x_n y_n$$

Dans cette expression, $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ sont les coordonnées, dans la base convenable, des vecteurs $\mathbf{x}$ et $\mathbf{y}$, et $p_1, p_2, \ldots, p_n$ sont des coefficients *positifs* choisis une fois pour toutes.

Exercice :

Vérifiez que dans $\mathbf{R}^3$, la formule :

$$f(\mathbf{x}, \mathbf{y}) = 3 x_1 y_1 + 2 x_2 y_2 + 5 x_3 y_3$$

définit bien une forme bilinéaire, symétrique et positive.

c) *Généralisation du produit scalaire : seconde définition*

Jusqu'ici, on a défini le produit scalaire de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ au moyen de leurs coordonnées par la règle de calcul :

$$\mathbf{x}.\mathbf{y} = x_1 y_1 + \ldots + x_n y_n$$

Cette règle est un cas particulier de la règle plus générale :

$$f(\mathbf{x}, \mathbf{y}) = p_1 x_1 y_1 + \ldots + p_n x_n y_n \qquad (\forall i, p_i > 0)$$

dans laquelle on calcule la somme des produits des coordonnées de même rang de x et de y pondérés par des coefficients positifs p_i.

On vient de voir que $f(\mathbf{x}, \mathbf{y})$ est une opération entre deux vecteurs qui jouit des propriétés fondamentales du produit scalaire (première définition) — bilinéarité, positivité, symétrie — et qu'inversement toute opération entre deux vecteurs jouissant de ces propriétés se calcule par cette formule plus générale.

On adopte donc comme *définition* (deuxième définition) générale d'un produit scalaire :

Un *produit scalaire* sur $\mathrm{V} = \mathbf{R}^n$ est une application $f : \mathrm{V} \times \mathrm{V} \to \mathbf{R}$ qui est :

— bilinéaire ;
— symétrique ;
— positive.

Cette définition équivaut à :

Le *produit scalaire* de deux vecteurs x et y se *calcule* par la règle :

$$f(\mathbf{x}, \mathbf{y}) = p_1 x_1 y_1 + p_2 x_2 y_2 + \ldots + p_n x_n y_n$$

où les coefficients p_i sont *positifs* et choisis une fois pour toutes.

En description statistique, les coefficients que l'on prend sont généralement *tous égaux entre eux*, et tous égaux :

— soit à 1 $\boxed{x_1 y_1 + \ldots + x_n y_n}$

— soit à $\frac{1}{n}$ $\boxed{\frac{x_1 y_1 + \ldots + x_n y_n}{n}}$ (cf. § 5 ci-dessous).

En calcul des probabilités (cf. chap. IX, § 3), ils sont en général *distincts*.

Comme les calculs et les écritures sont plus simples en prenant les coefficients tous égaux à 1, c'est ce qu'on va continuer à faire dans la suite de ce chapitre ; mais il doit être bien entendu que tous les résultats que l'on montrera dans ce cas particulier restent valables dans le cas général. La démonstration de ces résultats ne repose en effet que sur les trois propriétés de bilinéarité, symétrie et positivité.

3. Longueur d'un vecteur dans $\mathbf{R}^n$

a) *Longueur (norme) euclidienne d'un vecteur*

On a vu que pour tout vecteur $\mathbf{x}$ de $\mathbf{R}^n$:

$$\boxed{\mathbf{x}.\mathbf{x} = x_1^2 + x_2^2 + \ldots + x_n^2}$$

C'est la somme des carrés des coordonnées de $\mathbf{x}$; c'est un nombre toujours *positif ou nul*.

Par définition, on appelle *longueur euclidienne* (syn. : *norme* euclidienne) de $\mathbf{x}$ la racine carrée du nombre positif $\mathbf{x}.\mathbf{x}$. Cette longueur se note $\| \mathbf{x} \|$.

Longueur (norme) euclidienne de $\mathbf{x}$:

$$\| \mathbf{x} \| = (\mathbf{x}.\mathbf{x})^{1/2} = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$$

Exemple :

Dans $\mathbf{R}^4$, la longueur euclidienne de $\mathbf{x} = (2, -1, 1, -3)$ est :

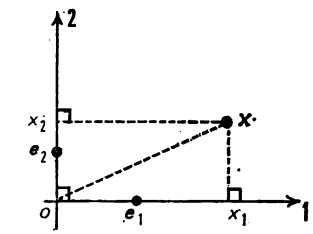
$$\| \mathbf{x} \| = (4 + 1 + 1 + 9)^{1/2} = 15^{1/2} \simeq 3{,}9$$

($\simeq$ signifie : approximativement égal à).

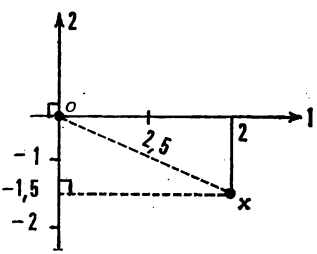
Cette définition de la longueur équivaut à :

> Le produit scalaire d'un vecteur $\mathbf{x}$ avec lui-même est le *carré de sa longueur* euclidienne :
>
> $$\mathbf{x}.\mathbf{x} = \| \mathbf{x} \|^2$$

Si l'on prend cette définition de la longueur d'un vecteur, et si on qualifie celle-ci d'euclidienne, c'est parce qu'elle généralise aux vectoriels de dimension quelconque le calcul des longueurs en géométrie (vectoriels de dimension 2 ou 3).



Considérons en effet le plan, rapporté à une base de deux vecteurs *orthogonaux* et de même longueur 1 ; alors, si $\mathbf{x} = (x_1, x_2)$ est un vecteur, la relation de Pythagore, exprime que la longueur du vecteur $\mathbf{x}$ est la racine carrée de $(x_1^2 + x_2^2)$: le carré de l'hypoténuse d'un triangle rectangle est la somme des carrés des côtés de l'angle droit.



Par exemple, le vecteur $\mathbf{x} = (2, -3/2)$ a pour longueur :

$$(4 + 9/4)^{1/2} = (25/4)^{1/2} = 5/2 = 2{,}5$$

Calculer la longueur des vecteurs $\mathbf{x}$ d'un vectoriel de dimension n par la formule : si $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, sa longueur est $(x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$ permet donc dans une certaine mesure de généraliser aux vectoriels de dimension n l'intuition que nous avons des longueurs (et des distances) en géométrie euclidienne.

N.B. — Jusqu'ici, les vectoriels que l'on considérait pouvaient indifféremment être sur $\mathbf{Q}$ ou sur $\mathbf{R}$; les coordonnées des vecteurs pouvaient indifféremment être des rationnels, ou des réels. C'est parce que toutes les propriétés mises en jeu, celles de la *linéarité*, ne faisaient appel qu'à la structure de corps ordonné de $\mathbf{Q}$ ou de $\mathbf{R}$ (cf. chap. I, § 4 et 5).

Dès l'instant où l'on veut calculer des longueurs euclidiennes, on doit *prendre des racines carrées :* or on a vu que seul le corps des réels contient toutes les racines carrées ; c'est pourquoi, chaque fois qu'il est question de longueurs, puis de distances et de métrique, et de calcul *algébrique* sur des longueurs, on ne *considère que des vectoriels sur* $\mathbf{R}$ (1).

b) *Formule* $(a + b)^2$ *pour les vecteurs*

Soit $\mathbf{x}$ et $\mathbf{y}$ deux vecteurs ; le produit scalaire de $(\mathbf{x} + \mathbf{y})$ avec lui-même est :

$$(\mathbf{x} + \mathbf{y}).(\mathbf{x} + \mathbf{y}) = \mathbf{x}.\mathbf{x} + \mathbf{x}.\mathbf{y} + \mathbf{y}.\mathbf{x} + \mathbf{y}.\mathbf{y}$$

Mais :

$$\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x} \qquad \text{et} \qquad \mathbf{x}.\mathbf{x} = \| \mathbf{x} \|^2 \qquad \mathbf{y}.\mathbf{y} = \| \mathbf{y} \|^2$$

$$(\mathbf{x} + \mathbf{y}).(\mathbf{x} + \mathbf{y}) = \| \mathbf{x} + \mathbf{y} \|^2$$

Donc :

$$\boxed{\| \mathbf{x} + \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + 2\,(\mathbf{x}.\mathbf{y}) + \| \mathbf{y} \|^2}$$

(1) Lorsqu'il s'agit de la métrique *euclidienne* ; on verra au chap. IV, § *4, qu'il est possible de définir des métriques pour lesquelles $\mathbf{Q}$ est suffisant.

C'est la formule : « $(a + b)^2 = a^2 + 2\,ab + b^2$ » pour les vecteurs.

Elle s'énonce :

Le carré de la longueur d'une somme de deux vecteurs est égal à la somme des carrés des longueurs de ces vecteurs et du double de leur produit scalaire.

4. Vecteurs orthogonaux

a) *Définition*

On a vu que :

— en général, le produit scalaire de deux vecteurs *n'est pas nul* ;
— si l'un des vecteurs est nul, le produit scalaire est nul ;
— mais le *produit scalaire peut être nul sans qu'aucun des deux vecteurs le soit* (voir exemples § 1-*b*).

Par définition, on dit que *deux vecteurs d'un même vectoriel* sont *orthogonaux* si et seulement si leur produit *scalaire vaut zéro.*

> $\mathbf{x}$ est *orthogonal* à $\mathbf{y}$ si et seulement si :
>
> $$\boxed{\mathbf{x}.\mathbf{y} = 0}$$
>
> *i.e.* : $x_1 y_1 + x_2 y_2 + \ldots + x_n y_n = 0.$

On note souvent l'orthogonalité de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ par le symbole $\perp$:

$\mathbf{x} \perp \mathbf{y}$ se lit $\mathbf{x}$ est orthogonal à $\mathbf{y}$.

Par définition : $\boxed{\mathbf{x} \perp \mathbf{y} \Leftrightarrow \mathbf{x}.\mathbf{y} = 0}$ (1).

(1) En particulier, tout vecteur est orthogonal au vecteur nul.

Exercice :

Le seul vecteur orthogonal à lui-même est le vecteur nul.

En effet : $\mathbf{x} \perp \mathbf{x}$ signifie $\mathbf{x}.\mathbf{x} = 0$.

On a vu ci-dessus : $\mathbf{x}.\mathbf{x} = \| \mathbf{x} \|^2$ (carré de la longueur de $\mathbf{x}$).

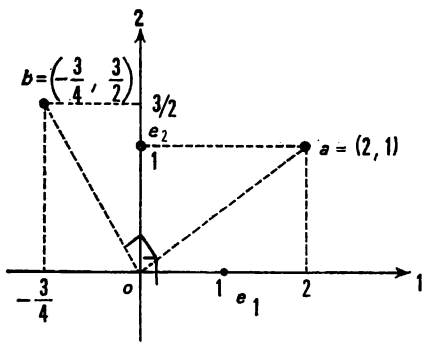
Or on a démontré au § 1-*d* que le seul vecteur dont la longueur est nulle est le vecteur nul :

$$\| \mathbf{x} \| = 0 \Rightarrow \mathbf{x} = \mathbf{0}$$

b) *Orthogonalité en géométrie et produit scalaire*

Pourquoi dit-on que deux vecteurs sont orthogonaux lorsque leur produit scalaire est nul ? Par extension de ce qui se passe en géométrie euclidienne dans le cas du vectoriel de dimension 2 ; considérons le plan où a été choisie une origine (vecteur nul **0**), et où la base est constituée de deux vecteurs *perpendiculaires* $\mathbf{e}_1$ et $\mathbf{e}_2$.

$$\mathbf{e}_1 = (1,0) \qquad \mathbf{e}_2 = (0,1)$$



On a : $\mathbf{e}_1.\mathbf{e}_2 = 1 \times 0 + 0 \times 1 = 0 + 0 = 0$.

De même, les vecteurs :

$$\mathbf{a} = (2,1) \qquad \text{et} \qquad \mathbf{b} = (-3/4, 3/2)$$

sont représentés par des points, des vecteurs d'origine **0**, perpendiculaires entre eux ; et l'on a :

$$\mathbf{a}.\mathbf{b} = 2 \times (-3/4) + 1 \times 3/2 = -3/2 + 3/2 = 0$$

D'une façon générale, on montre que deux vecteurs :

$$\mathbf{x} = (x_1, x_2) \qquad \mathbf{y} = (y_1 y_2)$$

sont orthogonaux (en géométrie plane) si et seulement si :

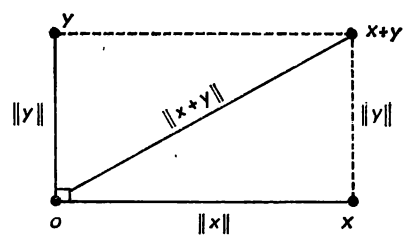
$$\mathbf{x}.\mathbf{y} = x_1 x_2 + y_1 y_2 = 0$$

c) *Relation de Pythagore*

La principale justification de la définition adoptée pour l'orthogonalité de deux vecteurs se trouve dans le théorème suivant :

Si **x** et **y** *sont deux vecteurs, le carré de la longueur de leur somme* $(\mathbf{x} + \mathbf{y})$ *est égal à la somme des carrés de leurs longueurs si et seulement si* **x** *et* **y** *sont orthogonaux.*

$$\mathbf{x} \perp \mathbf{y} \Leftrightarrow \| \mathbf{x} + \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + \| \mathbf{y} \|^2$$



En effet, on a (formule $(a + b)^2$), quels que soient $\mathbf{x}$ et $\mathbf{y}$:

$$\| \mathbf{x} + \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + 2 (\mathbf{x}.\mathbf{y}) + \| \mathbf{y} \|^2$$

Si $\mathbf{x} \perp \mathbf{y}$, alors $\mathbf{x}.\mathbf{y} = 0$, et il reste :

$$\| \mathbf{x} + \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + \| \mathbf{y} \|^2$$

Si réciproquement $\| \mathbf{x} + \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + \| \mathbf{y} \|^2$, alors $\| \mathbf{x} \|^2 + \| \mathbf{y} \|^2 = \| \mathbf{x} \|^2 + 2 (\mathbf{x}.\mathbf{y}) + \| \mathbf{y} \|^2$.

Donc : $\mathbf{x}.\mathbf{y} = 0$

Cette démonstration, qui ne suppose rien sur la dimension du vectoriel, montre que la définition prise pour l'orthogonalité ($\mathbf{x}.\mathbf{y} = 0$) est cohérente avec celle de la longueur ($\mathbf{x}.\mathbf{x} = \| \mathbf{x} \|^2$) et permet bien de généraliser ce qui se passe en géométrie plane euclidienne : le « théorème de Pythagore » est valable dans tout vectoriel sur $\mathbf{R}$.

Par conséquent, la figure ci-dessus de vecteurs orthogonaux peut s'interpréter, bien qu'elle soit dessinée à deux dimensions, dans un vectoriel quelconque.

5. Application a la description statistique : covariance, variance

a) *Covariance*

A chaque vecteur $\mathbf{x}$ de $\mathbf{R}^n$, le vectoriel de dimension n sur $\mathbf{R}$, on peut associer (*MSH*, chap. XIII) sa *moyenne arithmétique* $m_{\mathbf{x}}$ (parfois notée aussi $m(\mathbf{x})$).

$$m_{\mathbf{x}} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{S_{\mathbf{x}}}{n}$$

où l'on pose : $S_{\mathbf{x}} = x_1 + x_2 + \ldots + x_n$.

$S_{\mathbf{x}}$ est la *somme* (initiale S) des coordonnées de $\mathbf{x}$.

Le *vecteur moyen* $\mathbf{m}_x$ de $\mathbf{x}$ est celui dont toutes les coordonnées sont égales à sa moyenne arithmétique m_x :

$$\mathbf{m}_x = (m_x, m_x, \ldots, m_x)$$

On sait que la moyenne est une valeur centrale pour la suite des nombres $x_1, x_2, \ldots, x_n$, coordonnées de $\mathbf{x}$.

Si $\mathbf{y}$ désigne un second vecteur de $\mathbf{R}^n$, de vecteur moyen $\mathbf{m}_y$, on appelle *covariance* de $\mathbf{x}$ et de $\mathbf{y}$, notée Cov_{xy} (parfois $\mathrm{Cov}(\mathbf{x}, \mathbf{y})$) le produit scalaire du vecteur $(\mathbf{x} - \mathbf{m}_x)$ par le vecteur $(\mathbf{y} - \mathbf{m}_y)$, divisé par n.

$$\boxed{\mathrm{Cov}_{xy} = 1/n\,(\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y)}$$

On peut dire que la covariance de deux vecteurs est leur produit scalaire, chacun d'entre eux étant rapporté à sa moyenne, et en adoptant un produit scalaire avec une *pondération* $(p_1 x_1 y_1 + \ldots + p_n x_n y_n)$ dans laquelle tous les coefficients sont égaux (cf. § 2-*c*) à $1/n$.

La covariance, que l'on retrouvera en calcul des probabilités (II^e partie), intervient en description statistique principalement par le biais de la variance (§ 5-*c*) et du coefficient de corrélation linéaire (chap. V).

Exemple :

Dans $\mathbf{R}^3$, $\mathbf{x} = (3, -1, 1)$, $\mathbf{y} = (-2, 5, -4)$.

$$m_x = \frac{3 - 1 + 1}{3} = 1 \qquad m_y = \frac{-2 + 5 - 4}{3} = -\frac{1}{3}$$

$$\mathbf{x} - \mathbf{m}_x = (3 - 1, -1 - 1, 1 - 1) = (2, -2, 0)$$

$$\mathbf{y} - \mathbf{m}_y = (-2 - (-1/3), 5 - (-1/3), -4 - (-1/3)) = (-5/3, 16/3, -11/3)$$

$$(\mathbf{x}-\mathbf{m}_x).(\mathbf{y}-\mathbf{m}_y) = 2.(-5/3)+(-2).16/3+0.(-11/3)$$
$$= -10/3-32/3+0 = -42/3 = -14$$

$$\mathrm{Cov}_{xy} = \frac{(\mathbf{x}-\mathbf{m}_x).(\mathbf{y}-\mathbf{m}_y)}{n} = -\frac{14}{3} = -4{,}666...$$

b) *Calcul pratique de la covariance*

Selon les règles de calcul du produit scalaire, le produit $(\mathbf{x}-\mathbf{m}_x).(\mathbf{y}-\mathbf{m}_y)$ se développe en :

$$(\mathbf{x}-\mathbf{m}_x).(\mathbf{y}-\mathbf{m}_y) = \mathbf{x}.\mathbf{y}-\mathbf{m}_x.\mathbf{y}-\mathbf{m}_y.\mathbf{x}+\mathbf{m}_x.\mathbf{m}_y$$

Comme $\mathbf{m}_x$ et $\mathbf{m}_y$ ont chacun toutes leurs coordonnées égales respectivement à la moyenne m_x de $\mathbf{x}$ et à celle m_y de $\mathbf{y}$, on a : $\mathbf{x}.\mathbf{m}_y = x_1.m_y + x_2.m_y + \ldots + x_n.m_y = (x_1 + x_2 + \ldots + x_n)\, m_y = nm_x\, m_y$.

De même : $\mathbf{m}_x.\mathbf{y} = nm_x\, m_y$ et $\mathbf{m}_x.\mathbf{m}_y = nm_x\, m_y$; finalement :

$$(\mathbf{x}-\mathbf{m}_x).(\mathbf{y}-\mathbf{m}_y) = \mathbf{x}.\mathbf{y} - nm_x\, m_y$$

D'où :

$$\boxed{\mathrm{Cov}_{xy} = \frac{\mathbf{x}.\mathbf{y} - nm_x\, m_y}{n}} \quad \text{ou encore} \quad \boxed{\frac{\mathbf{x}.\mathbf{y}}{n} - m_x\, m_y}$$

En introduisant les notations S_x et S_y (somme des coordonnées), comme $S_x = nm_x$, $S_y = nm_y$, on peut encore écrire :

$$\boxed{\mathrm{Cov}_{xy} = \frac{1}{n}\left(\mathbf{x}.\mathbf{y} - \frac{S_x\, S_y}{n}\right)}$$

Pratiquement, c'est au moyen de cette dernière expression que l'on calcule le plus rapidement la covariance. Mais il n'est pas interdit de se servir de la première.

Exemple :

$$\mathbf{x} = (3, -1, 1) \qquad S_x = 3 - 1 + 1 = 3$$

$$\mathbf{y} = (-2, 5, -4) \qquad S_y = -2 + 5 - 4 = -1$$

$$\mathbf{x}.\mathbf{y} = 3 \times (-2) + (-1) \times 5 + 1 \times (-4) = -6 - 5 - 4 = -15$$

$$S_x S_y = 3 \times (-1) = -3$$

$$\mathrm{Cov}_{xy} = \frac{1}{3}\left(-15 - \frac{(-3)}{3}\right) = \frac{1}{3}(-15 + 1) = -\frac{14}{3}$$

c) *Variance*

En particulier, la covariance d'un vecteur avec lui-même est appelée sa *variance*, et notée V_x (parfois $V(\mathbf{x})$) :

$$\boxed{\mathrm{Cov}_{xx} = \frac{(\mathbf{x} - \mathbf{m}_x).(\mathbf{x} - \mathbf{m}_x)}{n} = V_x}$$

On verra (chap. IV) comment la variance est une *mesure de la dispersion* d'une suite de nombres. Pour l'instant, retenons que son calcul pratique se fait au moyen de l'expression qui sert pour la covariance, en y posant $\mathbf{x} = \mathbf{y}$:

$$V_x = \frac{\mathbf{x}.\mathbf{x} - n m_x . m_x}{n} = \frac{\| \mathbf{x} \|^2 - n.m_x^2}{n}$$

(On rappelle que $\| \mathbf{x} \|^2 = \mathbf{x}.\mathbf{x}$ est le carré de la longueur du vecteur $\mathbf{x}$.) Soit encore :

$$\boxed{V_x = \frac{\| \mathbf{x} \|^2}{n} - m_x^2}$$

$$\boxed{\begin{array}{c} n = \text{dimension} \\ \text{du vectoriel de } \mathbf{x} \end{array}}$$

Cette dernière expression est commode pour le calcul numérique de la variance d'un vecteur.

Voici ce que signifie chaque terme :

$$V_x = \frac{(\text{longueur de } x)^2}{\text{nombre des coordonnées de } x} - (\text{moyenne de } x)^2$$

Disposition des calculs

$$x = (x_1, x_2, x_3, x_4, \ldots, x_7)$$

x a sept coordonnées, il appartient au vectoriel $\mathbf{R}^7$ de dimension 7. On peut disposer les calculs de la façon suivante.

i	1 2 3 4 5 6 7		
x_i	x_1 x_2 x_3 x_4 x_5 x_6 x_7	Somme des x_i	$\frac{\text{Somme des } x_i}{7} = m_x$
x_i^2	x_1^2 x_2^2 x_3^2 ..... x_7^2	Somme des x_i^2	$\frac{\text{Somme des } x_i^2}{7} = \frac{\|x\|^2}{n}$

Exemple numérique :

$x = (2, -1, 2, 0, 3, 1, -2)$.

i	1 2 3 4 5 6 7	Sommes	
x_i	2 −1 2 0 3 1 −2	5	$m_x = \frac{5}{7} \simeq 0{,}73$
x_i^2	4 1 4 0 9 1 4	23	$\frac{\|x\|^2}{7} = \frac{23}{7} \simeq 3{,}23$

$$m_x^2 \simeq (0{,}73)^2 \simeq 0{,}53$$

$$V_x \simeq 3{,}23 - 0{,}53 = 2{,}7$$

— On peut, si l'on préfère, faire le calcul en introduisant l'expression S_x (somme des x_i) :

$$\boxed{V_x = 1/n\,(\|\,\mathbf{x}\,\|^2 - S_x^2/n)}$$

Dans l'exemple ci-dessus, le calcul donne alors :

$$V_x = \frac{1}{7}\left(23 - \frac{5^2}{7}\right) = \frac{1}{7}\left(23 - \frac{25}{7}\right)$$

$$= \frac{1}{7}\left(\frac{161 - 25}{7}\right) = \frac{136}{49} \simeq 2{,}7$$

d) *Variance d'une somme de vecteurs*

Soit maintenant à calculer la variance de la somme $\mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$ de deux vecteurs ; soit V_{y+x} cette variance. Par définition :

$$V_{x+y} = 1/n\,\|\,(\mathbf{x} + \mathbf{y}) - \mathbf{m}_{x+y}\,\|^2$$

$\mathbf{m}_{x+y}$ est le vecteur dont toutes les coordonnées sont égales à la moyenne arithmétique m_{x+y} de la somme des deux vecteurs ; on sait que :

$$m_{x+y} = m_x + m_y$$

car la moyenne est une application *linéaire*.

Donc le *vecteur* moyen de $(\mathbf{x} + \mathbf{y})$, $\mathbf{m}_{x+y}$, est la somme des vecteurs $\mathbf{m}_x$ et $\mathbf{m}_y$:

$$\mathbf{m}_{x+y} = \mathbf{m}_x + \mathbf{m}_y$$

Donc :

$$\| \mathbf{x} + \mathbf{y} - \mathbf{m}_{x+y} \|^2 = \| \mathbf{x} + \mathbf{y} - \mathbf{m}_x - \mathbf{m}_y \|^2$$
$$= \| (\mathbf{x} - \mathbf{m}_x) + (\mathbf{y} - \mathbf{m}_y) \|^2$$

Posons, pour alléger l'écriture :

$$\mathbf{x} - \mathbf{m}_x = \mathbf{u} \qquad \mathbf{y} - \mathbf{m}_y = \mathbf{v}$$

On a ainsi : $V_{x+y} = 1/n \| \mathbf{u} + \mathbf{v} \|^2$.

Or :

$$\| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + 2 (\mathbf{u}.\mathbf{v}) + \| \mathbf{v} \|^2 \qquad \text{(formule } (a+b)^2\text{)}$$

Donc :

$$V_{x+y} = \frac{1}{n} \| \mathbf{u} \|^2 + 2 \frac{(\mathbf{u}.\mathbf{v})}{n} + \frac{1}{n} \| \mathbf{v} \|^2$$

Or :

$$1/n \| \mathbf{u} \|^2 = 1/n \| \mathbf{x} - \mathbf{m}_x \|^2 = V_x$$
$$1/n \| \mathbf{v} \|^2 = 1/n \| \mathbf{y} - \mathbf{m}_y \|^2 = V_y$$
$$\frac{\mathbf{u}.\mathbf{v}}{n} = \frac{(\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y)}{n} = \mathrm{Cov}_{xy}$$

On a donc finalement :

$$\boxed{V_{x+y} = V_x + 2\,\mathrm{Cov}_{xy} + V_y}$$

Ce qui s'énonce : la variance d'*une somme* N'EST PAS en général la somme des variances ; la variance de la somme $\mathbf{x} + \mathbf{y}$ de deux vecteurs, c'est la somme des variances de $\mathbf{x}$ et de $\mathbf{y}$, et du double de la covariance de $\mathbf{x}$ et de $\mathbf{y}$.

N.B. — Cette égalité :

$$V_{x+y} = V_x + 2\,Cov_{xy} + V_y$$

n'est rien d'autre que la classique égalité (numérique) :

$$(a+b)^2 = a^2 + 2\,ab + b^2$$

Comme l'a fait apparaître la démonstration, c'est parce que cette égalité reste valable pour les vecteurs que l'on obtient l'expression ci-dessus de la variance d'une somme.

$$\begin{array}{rcl} \overbrace{V_{x+y}} & = & V_x + 2\,\overbrace{Cov_{xy}} + V_y \\ \overbrace{(a+b)^2} & = & \overbrace{a^2} + 2\,\overbrace{ab} + \overbrace{b^2} \end{array}$$

Exemple :

Soient les deux vecteurs :

$$\mathbf{x} = (3, -1, 1) \qquad \mathbf{y} = (-2, 5, -4)$$

On a : $m_x = 1$, $m_y = -1/3$, $Cov_{xy} = -14/3$.

$V_x = 8/3$ $\qquad V_y = 134/3$ $\qquad$ (faites le calcul)

D'où :

$$V_{x+y} = \frac{8 - 2\,(14) + 134}{3} = \frac{114}{3} = 38$$

Il y a néanmoins des cas où les variances s'ajoutent : c'est lorsque les vecteurs $(\mathbf{x} - \mathbf{m}_x)$ et $(\mathbf{y} - \mathbf{m}_y)$ sont *orthogonaux*, puisque cela signifie que leur produit scalaire est nul. On peut donc énoncer :

> La variance V_{x+y} d'une somme $\mathbf{x} + \mathbf{y}$ de deux vecteurs est la somme des variances de ces vecteurs si et seulement si les vecteurs $(\mathbf{x} - \mathbf{m}_x)$ et $(\mathbf{y} - \mathbf{m}_y)$ sont orthogonaux :
>
> $$V_{x+y} = V_x + V_y \Leftrightarrow (\mathbf{x} - \mathbf{m}_x) \perp (\mathbf{y} - \mathbf{m}_y)$$

Remarques :

1. On a vu que si deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ sont multipliés par un même nombre λ, leur produit scalaire est multiplié par λ^2 (§ 1-*c*, -*b*) ; comme le vecteur moyen de chacun d'entre eux est multiplié par λ, il en résulte :

$$\boxed{\mathrm{Cov}_{\lambda \mathbf{x},\, \lambda \mathbf{y}} = \lambda^2 \, \mathrm{Cov}_{\mathbf{xy}}}$$

Et en particulier :

$$\boxed{\mathrm{Var}_{\lambda \mathbf{x}} = \lambda^2 \, \mathrm{Var}_{\mathbf{x}}}$$

Multiplier un vecteur par un nombre λ, multiplie sa variance par le carré de λ.

2. La variance est le carré d'une longueur ; cette longueur elle-même, étudiée au chapitre IV, est l'*écart type* noté $\sigma_{\mathbf{x}}$ (sigma $\mathbf{x}$) du vecteur $\mathbf{x}$:

$$\sigma_{\mathbf{x}} = V_{\mathbf{x}}^{1/2}$$

6. Inégalité de Schwarz. Inégalité du triangle

a) *Inégalité de Schwarz*

Soit dans $\mathbf{R}^4$ les deux vecteurs :

$$\mathbf{x} = (-1, 3, -5, 1) \qquad \mathbf{y} = (-1, 3, 4, 0)$$

On a :

$$\| \mathbf{x} \| = [(-1)^2 + 3^2 + (-5)^2 + 1]^{1/2} = (36)^{1/2} = 6$$
$$\| \mathbf{y} \| = [(-1)^2 + 3^2 + 4^2 + 0]^{1/2} = (26)^{1/2} \simeq 5$$
$$\mathbf{x}.\mathbf{y} = +1 + 9 - 20 + 0 = -10$$

La valeur absolue du produit scalaire est :

$$| \mathbf{x}.\mathbf{y} | = | - 10 | = 10$$

Elle est inférieure au produit des longueurs des deux vecteurs égal approximativement à 30 :

$$|| \mathbf{x} || \times || \mathbf{y} || \simeq 6 \times 5 = 30$$

Il ne pouvait en être autrement ; *l'inégalité* dite de *Schwarz* que l'on va démontrer, prouve en effet que, quels que soient les vecteurs $\mathbf{x}$ et $\mathbf{y}$ d'un même vectoriel, la valeur absolue de leur produit scalaire est au plus égale au produit de leurs longueurs euclidiennes :

Inégalité de Schwarz :

$$| \mathbf{x}.\mathbf{y} | \leqslant || \mathbf{x} || \times || \mathbf{y} ||$$

De façon équivalente, l'inégalité de Schwarz peut encore s'écrire :

$$- || \mathbf{x} || \times || \mathbf{y} || \leqslant \mathbf{x}.\mathbf{y} \leqslant || \mathbf{x} || \times || \mathbf{y} ||$$

**Démonstration :*

L'inégalité est évidemment satisfaite lorsque $\mathbf{x} = \mathbf{0}$ ou $\mathbf{y} = \mathbf{0}$ car alors $\mathbf{x}.\mathbf{y} = 0$ et $|| \mathbf{x} || = 0$ ou $|| \mathbf{y} || = 0$.

Donc : $|| \mathbf{x} || \times || \mathbf{y} || = 0$.

Si $\mathbf{x} \neq \mathbf{0}$ et $\mathbf{y} \neq 0$ sont deux vecteurs d'un même vectoriel, considérons, λ étant un nombre quelconque, le carré :

$$|| \mathbf{x} - \lambda \mathbf{y} ||^2 = (\mathbf{x} - \lambda \mathbf{y}).(\mathbf{x} - \lambda \mathbf{y})$$

$\mathbf{x}$ et $\mathbf{y}$ étant donnés, quel que soit λ, ce carré est positif ou nul (puisque c'est le carré d'un nombre).

$$\forall\lambda \qquad \| \mathbf{x} - \lambda\mathbf{y} \|^2 \geqslant 0$$

Mais, l'on a :

$$(\mathbf{x} - \lambda\mathbf{y}).(\mathbf{x} - \lambda\mathbf{y}) = \| \mathbf{x} \|^2 - 2\lambda(\mathbf{x}.\mathbf{y}) + \lambda^2 \| \mathbf{y} \|^2$$

C'est là un « trinôme du second degré » en λ, qu'on peut encore écrire :

$$A\lambda^2 - 2B\lambda + C = A\left[\left(\frac{\lambda - B}{A}\right)^2 + \frac{AC - B^2}{A^2}\right]$$

En posant : $A = \| \mathbf{y} \|^2$, $B = \mathbf{x}.\mathbf{y}$, $C = \| \mathbf{x} \|^2$.

Ce trinôme doit être positif, quel que soit λ, donc en particulier pour $\lambda = B/A$.

Comme $A > 0$, on obtient ainsi :

$$AC - B^2 \geqslant 0 \quad \text{ou} \quad AC \geqslant B^2$$

C'est-à-dire :

$$\| \mathbf{x} \|^2 \times \| \mathbf{y} \|^2 \geqslant |\, \mathbf{x}.\mathbf{y} \,|^2$$

Soit, en prenant les racines carrées :

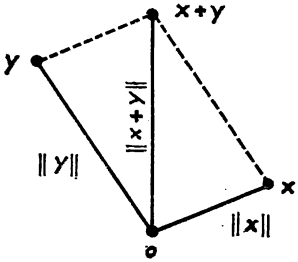
$$\| \mathbf{x} \| \times \| \mathbf{y} \| \geqslant |\, \mathbf{x}.\mathbf{y} \,|$$

b) *Application : inégalité du triangle*

La longueur $\| \mathbf{x} \|$ des vecteurs satisfaisait à l'*inégalité du triangle :*

$$(T) \qquad \boxed{\| \mathbf{x} + \mathbf{y} \| \leqslant \| \mathbf{x} \| + \| \mathbf{y} \|}$$

En géométrie euclidienne plane (vectoriel de dimension 2) cette inégalité est bien connue ; elle se dit encore : chaque



côté d'un triangle a une longueur au plus égale à la somme des longueurs des deux autres.

*L'inégalité de Schwarz permet de *prouver* l'inégalité du triangle pour un vectoriel de dimension quelconque sur **R**, et de voir ainsi que la définition prise pour la longueur d'un vecteur généralise bien cette propriété des longueurs de la géométrie.

En effet, l'inégalité (T) du triangle équivaut à :

$$|| \mathbf{x} + \mathbf{y} ||^2 \leqslant (|| \mathbf{x} || + || \mathbf{y} ||)^2$$

Or :

$$|| \mathbf{x} + \mathbf{y} ||^2 = (\mathbf{x} + \mathbf{y}).(\mathbf{x} + \mathbf{y}) = || \mathbf{x} ||^2 + 2\,(\mathbf{x}.\mathbf{y}) + || \mathbf{y} ||^2$$

et :

$$(|| \mathbf{x} || + || \mathbf{y} ||)^2 = || \mathbf{x} ||^2 + 2\,|| \mathbf{x} || \times || \mathbf{y} || + || \mathbf{y}^2 ||$$

Par conséquent, (T) équivaut à :

$$(T') \quad \mathbf{x}.\mathbf{y} \leqslant || \mathbf{x} || \times || \mathbf{y} ||$$

Comme tout nombre est au plus égal à sa valeur absolue, le nombre $\mathbf{x}.\mathbf{y}$ est au plus égal à sa valeur absolue $|\,\mathbf{x}.\mathbf{y}\,|$: $\mathbf{x}.\mathbf{y} \leqslant |\,\mathbf{x}.\mathbf{y}\,|$.

D'après l'inégalité de Schwarz :

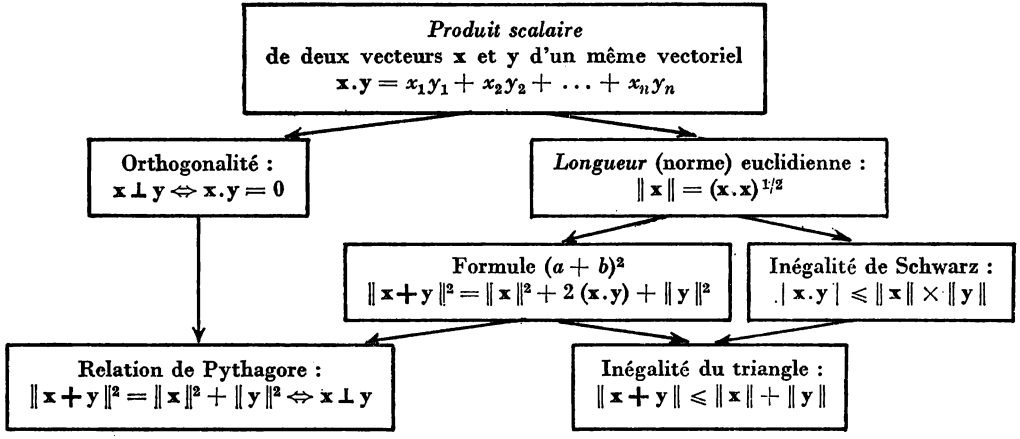
$$|\,\mathbf{x}.\mathbf{y}\,| \leqslant || \mathbf{x} || \times || \mathbf{y} ||$$

Donc *a fortiori* (transitivité de $\leqslant$) :

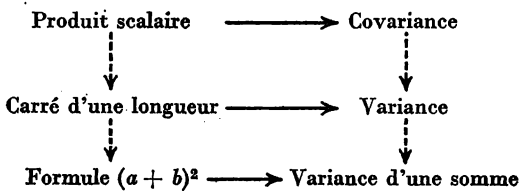
$$(T') \quad \mathbf{x}.\mathbf{y} \leqslant || \mathbf{x} || \times || \mathbf{y} ||$$

(T') équivaut à (T) ; donc (T) est prouvée.

Récapitulation


Produit scalaire
de deux vecteurs x et y d'un même vectoriel
$\mathbf{x}.\mathbf{y} = x_1y_1 + x_2y_2 + \ldots + x_ny_n$
Orthogonalité :
$\mathbf{x} \perp \mathbf{y} \Leftrightarrow \mathbf{x}.\mathbf{y} = 0$
Longueur (norme) euclidienne :
$\|\mathbf{x}\| = (\mathbf{x}.\mathbf{x})^{1/2}$
Formule $(a+b)^2$
$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2(\mathbf{x}.\mathbf{y}) + \|\mathbf{y}\|^2$
Inégalité de Schwarz :
$|\mathbf{x}.\mathbf{y}| \leqslant \|\mathbf{x}\| \times \|\mathbf{y}\|$
Relation de Pythagore :
$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \Leftrightarrow \mathbf{x} \perp \mathbf{y}$
Inégalité du triangle :
$\|\mathbf{x}+\mathbf{y}\| \leqslant \|\mathbf{x}\| + \|\mathbf{y}\|$

En statistique :



$$\mathrm{Cov}_{xy} = \frac{(\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y)}{n} = \frac{\mathbf{x}.\mathbf{y}}{n} - m_x m_y$$

$$V_x = \mathrm{Cov}_{xx} = \frac{\|\mathbf{x} - \mathbf{m}_x\|^2}{n} = \frac{\|\mathbf{x}\|^2}{n} - m_x^2$$

$$V_{x+y} = V_x + 2\,\mathrm{Cov}_{xy} + V_y$$

CHAPITRE IV

MESURES DE LA DISTANCE ET DE LA DISPERSION

1. LONGUEUR ET DISTANCE EUCLIDIENNES

a) *Mesurer les différences*

En description statistique, lorsque l'on étudie plusieurs vecteurs d'un même vectoriel V, il est souvent utile de pouvoir définir un écart, une distance, entre ceux-ci : au vu de trois suites de nombres, **a**, **b** et **c** par exemple, on se demande si **a** est plus « proche » que **b** de **c**.

Pour fixer les idées, supposons que **a**, **b** et **c** soient 3 vecteurs de $\mathbf{R}^5$:

$$\mathbf{a} = (-2, 1, 0, 2, 2)$$
$$\mathbf{b} = (2, 1, -1, 3, -1)$$
$$\mathbf{c} = (2, -1, 1, 1, 0)$$

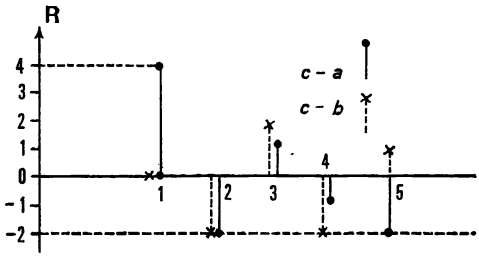
Une première idée naturelle : regarder en quoi diffèrent **c** et **a** d'une part, **c** et **b** d'autre part ; calculer les différences $\mathbf{c} - \mathbf{a}$ et $\mathbf{c} - \mathbf{b}$.

Rappelons comment est calculée la différence de deux vecteurs ; si $\mathbf{x} = (x_1, x_2, x_3, x_4, x_5)$ et $\mathbf{y} = (y_1, y_2, y_3, y_4, y_5)$, le vecteur $(\mathbf{x} - \mathbf{y})$ est par définition :

$$\mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2, x_3 - y_3, x_4 - y_4, x_5 - y_5)$$

Ici, on obtient :

$$\mathbf{c} - \mathbf{a} = (4, -2, 1, -1, -2)$$
$$\mathbf{c} - \mathbf{b} = (0, -2, 2, -2, 1)$$



La figure représente les graphiques « en bâtonnets » des deux différences $\mathbf{c} - \mathbf{a}$ et $\mathbf{c} - \mathbf{b}$.

Pour aller plus loin, et pouvoir décider sans ambiguïté si **a** est plus « éloigné » de **c** que **b**, il faut définir quelque chose qui joue le rôle de la longueur d'un vecteur, et de la distance de deux vecteurs.

Le calcul de cette « longueur » d'un vecteur peut se faire de diverses manières ; toutes prennent en compte l'ensemble des valeurs prises par les coordonnées du vecteur.

La plus usuelle de ces manières est celle qui est dans la logique de l'algèbre linéaire (l'algèbre des vectoriels) : prendre pour distance de deux vecteurs la *longueur euclidienne*, telle qu'elle a été définie (chap. III, § 3) au chapitre précédent. Au § 4 de ce chapitre-ci, on définira d'autres longueurs et distances qui, elles, ne relèveront plus aussi strictement de l'algèbre linéaire.

En ce qui concerne la longueur (et la distance) euclidienne, on a vu, au chapitre III, comment elles généralisent et étendent, à des vecteurs de dimension quelconque, certaines des propriétés essentielles des distances de la géométrie ; elles permettent donc de munir d'une *topologie* largement

analogue à celle de notre représentation de l'espace physique les listes et tableaux de nombres de la statistique.

Ci-dessous, on rappelle les définitions et propriétés principales.

b) *Longueur et distance euclidiennes*

— Dans le vectoriel $\mathbf{R}^n$ des suites de n nombres réels, on appelle *longueur* euclidienne (syn. : *norme euclidienne*) d'un vecteur $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ la racine carrée *positive* de la somme des carrés de ses coordonnées ; la longueur de $\mathbf{x}$ est notée $\| \mathbf{x} \|$.

Longueur (norme) *euclidienne* de $\mathbf{x}$:
$$\| \mathbf{x} \| = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$$

— Dans $\mathbf{R}^n$, la *distance* euclidienne de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ est par définition la *longueur euclidienne de leur différence* ; elle se note par conséquent $\| \mathbf{x} - \mathbf{y} \|$.

La différence $(\mathbf{x} - \mathbf{y})$ de $\mathbf{x}$ et $\mathbf{y}$ a pour coordonnées :

$$\mathbf{x} - \mathbf{y} = (x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n)$$

On a donc :

$$\| \mathbf{x} - \mathbf{y} \| = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2]^{1/2}$$

Distance euclidienne de $\mathbf{x}$ et $\mathbf{y}$: $\| \mathbf{x} - \mathbf{y} \| = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \ldots + (x_n - y_n)^2]^{1/2}$

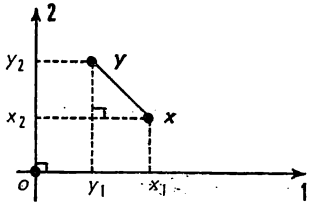
Exemple :

Pour les vecteurs $\mathbf{a}$, $\mathbf{b}$ et $\mathbf{c}$ du § 1-*a*, on a :

$$\| \mathbf{c} - \mathbf{a} \| = (16 + 4 + 1 + 1 + 4)^{1/2} = 26^{1/2} \simeq 5{,}1$$
$$\| \mathbf{c} - \mathbf{b} \| = (0 + 4 + 4 + 4 + 1)^{1/2} = 13^{1/2} \simeq 3{,}6$$

On conclut que $\mathbf{a}$ est moins proche de $\mathbf{b}$ que $\mathbf{c}$.

— Dans $\mathbf{R}^2$, représenté par le plan euclidien où l'on a choisi une origine **0** et deux axes rectangulaires, $\| \mathbf{x} - \mathbf{y} \|$, est bien la distance entre **x** et **y** : longueur du segment joignant **x** à **y** ; car, d'après la relation de Pythagore pour les triangles rectangles, cette longueur (hypoténuse) est : $((x_1 - y_1)^2 + (x_2 - y_2)^2)^{1/2}$.



— On rappelle que : $\| \mathbf{x} \|^2 = \mathbf{x}.\mathbf{x}$. Le carré de la longueur d'un vecteur **x** est le produit scalaire de **x** avec lui-même.

c) *Propriétés de la longueur euclidienne*

Associer à chaque vecteur **x** de $\mathbf{R}^n$ *sa* longueur, c'est définir une *application* l de $\mathbf{R}^n$ dans l'ensemble des réels positifs ou nuls, noté ici $\mathbf{R}_0^+$.

$$l : \mathbf{R}^n \to \mathbf{R}_0^+$$

Dans cette application l, *l'image* $l(\mathbf{x})$ d'un vecteur **x** est par définition sa longueur :

$$l(\mathbf{x}) = \| \mathbf{x} \|$$

C'est donc toujours un nombre positif ou nul :

1. $$\boxed{l(\mathbf{x}) = \| \mathbf{x} \| \geqslant 0 \quad (\| \mathbf{x} \| \in \mathbf{R}_0^+)}$$

Mais $\| \mathbf{x} \|$ n'est nul que si **x** est le vecteur nul, comme on l'a démontré (chap. III, § 3).

1 *bis.* La longueur $\| \mathbf{x} \|$ d'un vecteur est *nulle* si et seulement si ce vecteur est le *vecteur nul :*

$$\| \mathbf{x} \| = 0 \Leftrightarrow \mathbf{x} = \mathbf{0} = (0, 0, \ldots, 0)$$

1 et 1 *bis* constituent la propriété de *positivité* de la longueur.

— *Homothéties :*

Que peut-on dire de la longueur du produit d'un vecteur $\mathbf{x}$ par un nombre λ (homothétique de $\mathbf{x}$ dans le rapport λ) ?

$$\lambda\mathbf{x} = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$$

On a vu (chap. III, § 1) que pour le produit scalaire :

$$(\lambda\mathbf{x}).(\lambda\mathbf{y}) = \lambda^2 (\mathbf{x}.\mathbf{y})$$

En particulier :

$$\| \lambda\mathbf{x} \|^2 = (\lambda\mathbf{x}).(\lambda\mathbf{x}) = \lambda^2 (\mathbf{x}.\mathbf{x}) = \lambda^2 \| \mathbf{x} \|^2$$

Donc :

$$\| \lambda\mathbf{x} \| = (\lambda^2)^{1/2} \| \mathbf{x} \|$$

Or, $(\lambda^2)^{1/2}$ est la racine carrée *positive* de λ^2 ; c'est donc la *valeur absolue* de λ, notée $| \lambda |$. Finalement :

$$\| \lambda\mathbf{x} \| = | \lambda | \times \| \mathbf{x} \|$$

Dans $\mathbf{R}^4$, soit $\mathbf{x} = (2, -1, -2, 0)$.

$$\| \mathbf{x} \| = (4 + 1 + 4 + 0)^{1/2} = 9^{1/2} = 3$$

Prenons $\lambda = -3/2$.

$\| \lambda\mathbf{x} \|$ se calcule sans avoir besoin de passer par l'intermédiaire du calcul des coordonnées du vecteur $\lambda\mathbf{x}$.

$$\| \lambda\mathbf{x} \| = | \lambda | \times \| \mathbf{x} \| = | -3/2 | \times 3$$
$$= 3/2 \times 3 = 9/2 = 4{,}5$$

Cas particulier :

Tout vecteur $\mathbf{x}$ a même longueur que son opposé $-\mathbf{x}$:

$$\| -\mathbf{x} \| = \| (-1).\mathbf{x} \| = 1 \times \| \mathbf{x} \| = \| \mathbf{x} \|$$

2. La longueur euclidienne du produit par le nombre λ du vecteur $\mathbf{x}$ est le produit de la longueur de $\mathbf{x}$ par la valeur absolue de λ :

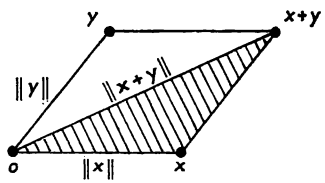
$$\| \lambda\mathbf{x} \| = | \lambda | \, \| \mathbf{x} \|$$

En particulier :

2 *bis.* Deux vecteurs *opposés* ont même longueur :

$$\| -\mathbf{x} \| = \| \mathbf{x} \|$$

Inégalité du triangle :



Que peut-on dire enfin de la longueur d'une somme de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$? Dans le cas des vecteurs de la géométrie euclidienne, pour deux vecteurs $\mathbf{x}$, $\mathbf{y}$ et leur somme $\mathbf{x} + \mathbf{y}$ construite selon la règle du parallélogramme, on a, d'après l'inégalité « tout côté d'un triangle a une longueur au plus égale à la somme des longueurs des deux autres côtés » appliquée au triangle hachuré sur la figure :

$$\| \mathbf{x} + \mathbf{y} \| \leqslant \| \mathbf{x} \| + \| \mathbf{y} \|$$

On a vu qu'une conséquence de l'inégalité de Schwarz est que, *quelle que soit la dimension du vectoriel* $\mathbf{R}^n$, *cette inégalité reste valable* (chap. III, § 6).

3. *Inégalité du triangle :* La longueur de la somme de deux vecteurs est au plus égale à la somme des longueurs de ces vecteurs :

$$\| x + y \| \leqslant \| x \| + \| y \|$$

Exemple :

Soit, dans $\mathbf{R}^4$:

$x = (2, -1, -2, 0)$ $\quad \| x \| = 3$ (faites le calcul)
$y = (1, 2, 4, -2)$ $\quad \| y \| = 5$

$x + y = (3, 1, 2, -2)$ $\quad \| x + y \| \simeq 4{,}2$ $(18^{1/2})$

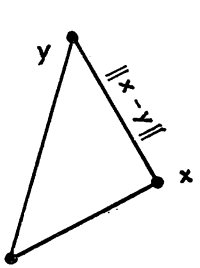
$$\| x + y \| \simeq 4{,}2 < 3 + 5 = 8 = \| x \| + \| y \|$$

N.B. — Grâce à cette inégalité, on peut affirmer que si deux vecteurs sont « petits » (si leurs longueurs sont petites), leur somme est elle aussi petite ; par exemple, si la longueur des vecteurs x et y respectivement est inférieure à 0,01, la longueur de leur somme ne dépasse pas 0,02 :

$$\| x + y \| \leqslant \| x \| + \| y \| \leqslant 0{,}01 + 0{,}01 = 0{,}02$$

d) *Propriétés des distances euclidiennes*

Puisque la distance euclidienne de deux vecteurs d'un même vectoriel est la longueur $\| x - y \|$ de leur différence, chacune des 3 propriétés de la norme (longueur) peut se traduire en une propriété correspondante de la distance.



Tout d'abord, la distance fait correspondre à *chaque couple* (x, y) de vecteurs du vectoriel $V = \mathbf{R}^n$ un nombre réel positif, et un seul ; c'est donc une application d du produit cartésien $V \times V$ de V par lui-

même ($V \times V$ est l'ensemble des couples $(\mathbf{x}, \mathbf{y})$, $\mathbf{x} \in V$, $\mathbf{y} \in V$) dans $\mathbf{R}_0^*$, ensemble des réels positifs ou nuls :

$$d : V \times V \to \mathbf{R}_0^*$$

L'image d'un couple $(\mathbf{x}, \mathbf{y})$ de vecteurs est la distance de $\mathbf{x}$ et de $\mathbf{y}$:

$$d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \|$$

D'après la première propriété des longueurs, on a (1) :

1. $$\boxed{d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \| \geqslant 0}$$

En outre, $\| \mathbf{x} - \mathbf{y} \| = 0$ équivaut (propriété 1 *bis* des longueurs) à $\mathbf{x} - \mathbf{y} = \mathbf{0}$, donc à :

$$\mathbf{x} = \mathbf{y}$$

1 *bis*. La distance euclidienne de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ est nulle si et seulement si $\mathbf{x} = \mathbf{y}$:

$$d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \| = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$$

Cette propriété était désirable pour une « distance » ; la distance exprime le degré de proximité entre deux vecteurs, par exemple entre deux séries statistiques ; qu'elle ne s'annule que lorsque les deux vecteurs sont confondus, identiques, permet de dire que si la distance de deux vecteurs est positive, ceux-ci sont certainement distincts.

La seconde propriété des longueurs donne immédiatement :

2. Si deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ sont multipliés par un même

(1) Cette propriété n'est qu'une autre expression du fait que $d(x, y) \in \mathbf{R}_0^*$.

nombre λ, leur distance euclidienne est multipliée par la valeur absolue de λ :

$$d(\lambda\mathbf{x}, \lambda\mathbf{y}) = \| \lambda\mathbf{x} - \lambda\mathbf{y} \| = \| \lambda (\mathbf{x} - \mathbf{y}) \|$$
$$= | \lambda | \, \| \mathbf{x} - \mathbf{y} \| = | \lambda | \, d(\mathbf{x}, \mathbf{y})$$

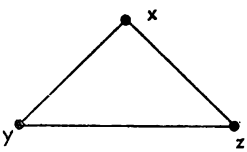
$$\boxed{d(\lambda\mathbf{x}, \lambda\mathbf{y}) = | \lambda | \, d(\mathbf{x}, \mathbf{y})}$$

En *particulier*, les deux vecteurs $\mathbf{x} - \mathbf{y}$ et $\mathbf{y} - \mathbf{x}$ qui sont *opposés* ont même longueur ; donc :

2 *bis.* $$\boxed{d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})}$$

La distance entre deux vecteurs ne dépend pas de l'ordre dans lequel on les prend. On dit qu'elle est *symétrique* par rapport à $\mathbf{x}$ et $\mathbf{y}$.

Enfin, *l'inégalité du triangle* reste vraie pour les distances ; si $\mathbf{x}$, $\mathbf{y}$ et $\mathbf{z}$ sont 3 vecteurs de $\mathbf{R}^n$, la distance entre $\mathbf{x}$ et $\mathbf{y}$ est au plus égale à la somme des distances entre $\mathbf{x}$ et $\mathbf{z}$ d'une part, entre $\mathbf{y}$ et $\mathbf{z}$ d'autre part.



$$d(\mathbf{x}, \mathbf{y}) \leqslant d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$$

En effet :

$$d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \|$$
$$d(\mathbf{x}, \mathbf{z}) = \| \mathbf{x} - \mathbf{z} \|$$
$$d(\mathbf{z}, \mathbf{y}) = \| \mathbf{y} - \mathbf{z} \|$$

Or on a : $x - y = (x - z) + (z - y)$.

Donc, d'après l'inégalité du triangle pour les longueurs :

$$\| x - y \| \leqslant \| x - z \| + \| z - y \|$$

3. *Inégalité du triangle :* Étant donné trois vecteurs x, y et z de $\mathbf{R}^n$, la distance euclidienne de deux quelconques d'entre eux est au plus égale à la somme de leurs distances au troisième :

$$d(x, y) \leqslant d(x, z) + d(z, y)$$

Exemple :

$x = (1, 0, 0, -1)$ $\quad y = (-1, 0, 1, 0)$ $\quad z = (2, -2, 1, 0)$

$x - y = (2, 0, -1, 1)$ $\quad d(x, y) = (4+0+1+1)^{1/2} = \sqrt{6}$

$x - z = (-1, 2, -1, 1)$ $\quad d(x, z) = (1+4+1+1)^{1/2} = \sqrt{7}$

$z - y = (3, -2, 0, 0)$ $\quad d(y, z) = (9+4+0+0)^{1/2} = \sqrt{13}$

On a évidemment :

$$\sqrt{6} < \sqrt{7} + \sqrt{13} \quad \text{et} \quad \sqrt{7} < \sqrt{6} + \sqrt{13}$$

On a enfin bien :

$$\sqrt{13} < \sqrt{7} + \sqrt{6}$$

Vérification :

$$(\sqrt{13})^2 = 13 < (\sqrt{7} + \sqrt{6})^2 = 7 + 6 + 2\sqrt{7}\sqrt{6} = 13 + 2\sqrt{7}\sqrt{6}$$

N.B. — L'inégalité du triangle permet de dire que si deux vecteurs sont proches d'un même troisième, ils sont proches l'un de l'autre.

c) *Généralisation du mode de calcul des distances euclidiennes*

La distance euclidienne étant définie à partir du produit scalaire, de même que l'on peut donner une définition plus générale de celui-ci au moyen d'une pondération des coordonnées (§ 2-*c*, chap. III) :

$$\mathbf{x}.\mathbf{y} = p_1 x_1 y_1 + p_2 x_2 y_2 + \ldots + p_n x_n y_n$$

de même, on peut définir une longueur et une distance, qui continuent à être qualifiées d'euclidiennes, par les règles :

$$\| \mathbf{x} \|^2 = p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2$$
$$l(\mathbf{x}) = \| \mathbf{x} \| = (p_1 x_1^2 + p_2 x_2^2 + \ldots + p_n x_n^2)^{1/2}$$
$$d(\mathbf{x}, \mathbf{y}) = \| \mathbf{x} - \mathbf{y} \| = [p_1 (x_1 - y_1)^2 + \ldots + p_n (x_n - y_n)^2]^{1/2}$$

Dans ces expressions, les coefficients p_i sont *tous positifs* (1) :

$$\forall i, p_i > 0$$

En description statistique, les pondérations utilisées sont le plus souvent des pondérations *uniformes* :

$$-p_1 = p_2 = \ldots = p_n = 1$$
$$-p_1 = p_2 = \ldots = p_n = 1/n$$

En calcul des probabilités, on a au contraire souvent recours à d'autres pondérations, toutes telles que la somme des poids vaut 1 ; on dit parfois dans ce cas « écart moyen quadratique », à la place de « distance euclidienne ».

Toutes les propriétés des longueurs et des distances euclidiennes énoncées dans ce chapitre, où l'on prend des *pondérations uniformes*, sont valables dans le cas général.

(1) Si certains coefficients p_i étaient nuls, on pourrait avoir $\| \mathbf{x} \| = 0$ sans que $\mathbf{x}$ soit nul.

2. Moyenne et distance minimum

a) *Un exemple*

On connaît (*MSH*, chap. XIII) des raisons pour lesquelles la *moyenne* arithmétique : $m_{\mathbf{x}} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ des coordonnées d'un vecteur $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ peut être considérée comme un « bon résumé » de $\mathbf{x}$, lorsque le vecteur $\mathbf{x}$ est *une statistique*, une suite de nombres, c'est-à-dire une application d'un ensemble $E = \{1, 2, \ldots, n\}$ dans $\mathbf{R}$. Ces raisons sont :

— *la linéarité :* la moyenne d'une somme est la somme des moyennes ; multiplier un vecteur par un nombre multiplie sa moyenne par ce nombre ;

— *la positivité :* si $\mathbf{x}$ est « plus grand » que $\mathbf{y}$ en ce sens que chaque coordonnée de $\mathbf{x}$ est plus grande que la coordonnée correspondante de $\mathbf{y}$ $(\forall_i, x_i \geqslant y_i)$, la moyenne de $\mathbf{x}$ est supérieure à celle de $\mathbf{y}$;

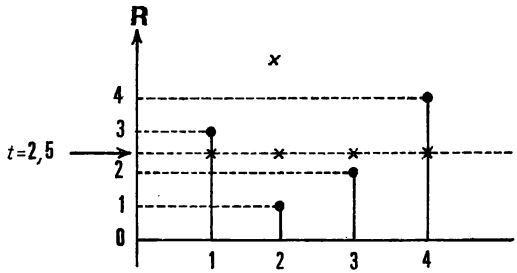
— si toutes les coordonnées de $\mathbf{x}$ sont égales entre elles, sa moyenne est leur valeur commune ;

— enfin, la moyenne arithmétique ne dépend pas de l'ordre dans lequel sont écrites les coordonnées.

Voici une raison supplémentaire de ce choix, qui fait comprendre en quoi la moyenne arithmétique d'une suite de nombres est une *valeur centrale* de cette suite.

Soit par exemple le vecteur $\mathbf{x} = (3, 1, 2, 4)$ de $\mathbf{R}^4$. C'est une application de $E = \{1, 2, 3, 4\}$ dans $\mathbf{R}$. $\mathbf{x} : E \to \mathbf{R}$.

$$x_1 = x(1) = \text{image de } 1 = 3$$
$$x_2 = x(2) = 1$$
$$x_3 = x(3) = 2$$
$$x_4 = x(4) = 4$$



Quel est, parmi les vecteurs **t** de $\mathbf{R}^4$ dont toutes les coordonnées sont égales, c'est-à-dire encore parmi les applications constantes de E dans **R** ($\mathbf{t} = (t, t, t, t)$), celui qui est le *plus proche* de **x** ? Ce vecteur **t** constituera la *meilleure approximation* de **x** par un vecteur constant (la meilleure approximation de l'application **x** par une application constante, de la suite des nombres 3, 1, 2, 4 par une suite constante).

Par exemple, $\mathbf{x} = (3, 1, 2, 4)$ peut représenter les notes obtenues par 4 sujets à un même test ; si tous avaient eu la *même note*, quelle est la meilleure façon de choisir cette note uniforme, celle qui est la plus proche des notes effectivements constatées ?

On va donc calculer t de façon que le vecteur **t** soit à *distance minimum* de **x**. Or, pour que $d(\mathbf{t}, \mathbf{x})$ soit minimum, il faut et suffit que son carré soit minimum.

$$d^2(\mathbf{t}, \mathbf{x}) = \| \mathbf{x} - \mathbf{t} \|^2$$
$$= (3 - t)^2 + (1 - t)^2 + (2 - t)^2 + (4 - t)^2$$

Développons chaque carré selon la « formule » :

$$(a + b)^2 = a^2 + 2ab + b^2$$

$$\| \mathbf{x} - \mathbf{t} \|^2 = (9 - 6t + t^2) + (1 - 2t + t^2) + (4 - 4t + t^2) + (16 - 8t + t^2)$$
$$= (9 + 1 + 4 + 16) - 2(3 + 1 + 2 + 4)t + 4t^2$$
$$= 30 - 2 \times 10t + 4t^2$$

C'est un trinôme du second degré ; il s'écrit encore :

$$\| \mathbf{x} - \mathbf{t} \|^2 = 4 \left(t^2 - 2 \times \frac{10}{4} t + \frac{30}{4} \right)$$
$$= 4 \left[\left(t - \frac{10}{4} \right)^2 + \frac{30}{4} - \frac{100}{16} \right]$$

La valeur de t qui rend l'expression minimum est celle qui annule $(t - 10/4)^2$, c'est-à-dire :

$$t = \frac{10}{4} = \frac{3 + 1 + 2 + 4}{4} = 2{,}5$$

C'est la moyenne arithmétique des coordonnées de $\mathbf{x}$; la distance minimum obtenue a d'ailleurs pour carré :

$$4 \left[\frac{30}{4} - \frac{100}{16} \right] = 30 - \frac{100}{4} = 5$$

Ainsi, la meilleure approximation de $\mathbf{x}$ par une constante est la moyenne arithmétique de ses coordonnées, ici 2,5, et la distance est $\sqrt{5}$.

b) *Cas général*

On suppose maintenant que l'on est dans le vectoriel $\mathbf{R}^n$. On *donne :*

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

On *cherche :* $\mathbf{t} = (t, t, t, \ldots, t)$ tel que $d(\mathbf{x}, \mathbf{t})$ soit minimum. En appliquant la formule $(a + b)^2$ pour les vecteurs (chap. III, § 3-*b*) il vient, selon un même calcul formel qu'au § 5-*b*, chap. III :

$$\| \mathbf{x} - \mathbf{t} \|^2 = (\mathbf{x} - \mathbf{t}).(\mathbf{x} - \mathbf{t}) = \| \mathbf{x} \|^2 - 2 (\mathbf{t}.\mathbf{x}) + \| \mathbf{t} \|^2$$
$$= \| x \|^2 - 2 n t m_{\mathbf{x}} + n t^2$$
$$= n \left[(t - m_{\mathbf{x}})^2 + \| \mathbf{x} \|^2 / n - m_{\mathbf{x}}^2 \right]$$

où $m_{\mathbf{x}} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$.

D'autre part, on se souvient que (chap. III, § 5-*c*) :

$$\frac{\| \mathbf{x} \|^2}{n} - m_{\mathbf{x}}^2 = \mathrm{V}_{\mathbf{x}} \qquad \text{(variance de } \mathbf{x})$$

Donc :

$$\boxed{d(\mathbf{x}, \mathbf{t})^2 = \| \mathbf{x} - \mathbf{t} \|^2 = n\,[(t - m_{\mathbf{x}})^2 + \mathrm{V}_{\mathbf{x}}]}$$

Pour obtenir le minimum, il faut prendre :

$$t = m_{\mathbf{x}} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

La distance *minimum* obtenue a alors pour carré *n fois la variance de* **x**.

$$\boxed{d^2(\mathbf{x}, \mathbf{m}_{\mathbf{x}}) = n\mathrm{V}_{\mathbf{x}}} \qquad (= \| \mathbf{x} \|^2 - n m_{\mathbf{x}}^2)$$

où : $\mathbf{m}_{\mathbf{x}} = (m_{\mathbf{x}}, m_{\mathbf{x}}, \ldots, m_{\mathbf{x}})$ est le *vecteur moyen* de **x**.

Cette distance minimum est donc elle-même égale à :

$$\boxed{d(\mathbf{x}, \mathbf{m}_{\mathbf{x}}) = (n\mathrm{V}_{\mathbf{x}})^{1/2}}$$

Exemple :

$$\mathbf{x} = (-5, 2, 7, 4, 6, 1) \qquad (\mathbf{x} \in \mathbf{R}^6)$$

$$m_{\mathbf{x}} = \frac{-5 + 2 + 7 + 4 + 6 - 1}{6} = \frac{13}{6} = 2{,}666\ldots$$

$$\| \mathbf{x} \|^2 = 25 + 4 + 49 + 16 + 36 + 1 = 131$$

$$n.m_{\mathbf{x}}^2 = 6 \cdot \frac{13^2}{6^2} = \frac{13^2}{6} = \frac{169}{6} \simeq 28$$

$$(\| \mathbf{x} \|^2 - n.m_{\mathbf{x}}^2)^{1/2} \simeq (131 - 28)^{1/2} = (104)^{1/2} \simeq 10$$

Le vecteur constant (dont toutes les coordonnées sont égales entre elles) $\mathbf{t} = (t, t, t, \ldots, t)$ à distance minimum d'un vecteur $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ *donné*, est celui pour lequel :

$$t = m_x = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

$$= \text{moyenne arithmétique de } \mathbf{x}$$

C'est le vecteur moyen de **x**.
La distance minimum est :

$$(\| \mathbf{x} \|^2 - n.m_{\mathbf{x}}^2)^{1/2} = (n\mathrm{V}_{\mathbf{x}})^{1/2}$$

où $\mathrm{V}_{\mathbf{x}}$ est la variance **x**.

N.B. — Si la distance était définie avec une pondération des coordonnées :

$$d(\mathbf{x}, \mathbf{y}) = (p_1 (x_1 - y_1)^2 + \ldots + p_n (x_n - y_n)^2)^{1/2}$$

avec $\forall_i, p_i > 0$, et si la somme des coefficients p_i est égale à 1 :

$$p_1 + p_2 + \ldots + p_n = 1$$

comme c'est fréquent en calcul des probabilités, la valeur centrale à distance minimum est la *moyenne pondérée* :

$$m_{\mathbf{x}} = p_1 x_1 + p_2 x_2 + \ldots + p_n x_n$$

La distance minimum a alors pour carré la variance $\mathrm{V}\mathbf{x}$, égale dans ce cas à :

$$\mathrm{V}_{\mathbf{x}} = p_1 (x_1 - m_{\mathbf{x}})^2 + p_2 (x_2 - m_{\mathbf{x}})^2 + \ldots + p_n (x_n - m_{\mathbf{x}})^2$$

3. Mesure de la dispersion : variance et écart type

On vient d'établir que le carré de la distance d'un vecteur $\mathbf{x} \in \mathbf{R}^n$ à son vecteur moyen $\mathbf{m}_x$ s'exprime au moyen de sa variance par :

$$d^2(\mathbf{x}, \mathbf{m}_x) = n\mathrm{V}_x$$

Ceci s'écrit encore :

$$\mathrm{V}_x = \frac{1}{n} d^2(\mathbf{x}, \mathbf{m}_x) = \frac{\| \mathbf{x} - \mathbf{m}_x \|^2}{n}$$

La variance s'interprète donc comme le carré de la distance de $\mathbf{x}$ à son vecteur $\mathbf{m}_x$ divisé par la dimension n de $\mathbf{x}$.

On a explicitement, en fonction des coordonnées $x_1, x_2, \ldots, x_n$ de $\mathbf{x}$:

$$\mathrm{V}_x = \frac{(x_1 - m_x)^2 + (x_2 - m_x)^2 + \ldots + (x_n - m_x)^2}{n}$$

Sur cette expression, on voit que la variance est la moyenne arithmétique des *carrés des écarts* $(x_i - m_x)$ des coordonnées de $\mathbf{x}$ à la moyenne de $\mathbf{x}$. Les écarts peuvent être positifs ou négatifs ; prendre leurs carrés consiste à ce que tous fournissent une contribution *positive* au calcul de la variance.

Plus ces écarts sont grands, plus la variance est grande : autrement dit, la variance est une moyenne des *fluctuations* de la suite des nombres $x_1, x_2, \ldots, x_n$ autour de leur moyenne m_x. Elle *mesure la dispersion* de cette suite de nombres (1).

Pour que ce point soit bien clair, considérons deux vecteurs de $\mathbf{R}^6$:

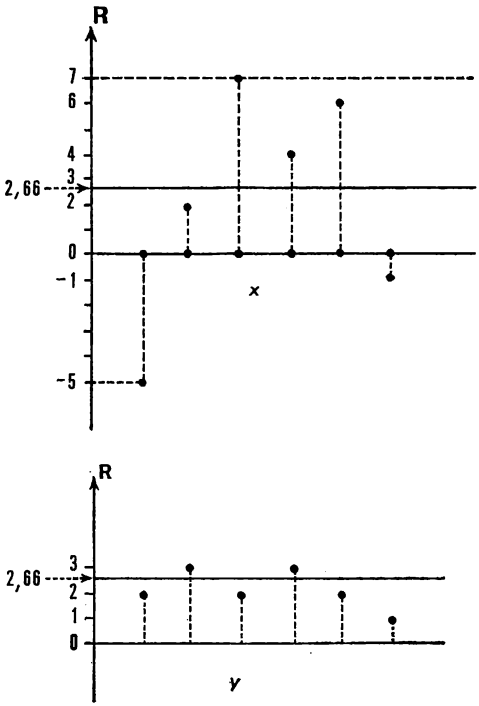
$$\mathbf{x} = (-5, 2, 7, 4, 6, -1)$$
$$\mathbf{y} = (2, 3, 2, 3, 2, 1)$$

(1) En particulier, la dispersion est nulle en ce sens que $\mathrm{V}_x = 0$ si et seulement si $x_1 = x_2 = \ldots = x_n$.

Ils ont pour moyennes : $m_x = 13/6 = 2{,}666...$

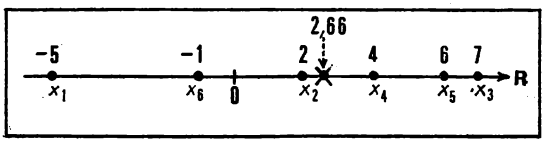
$$m_y = \frac{2+3+2+3+2+1}{6} = \frac{13}{6} = 2{,}66...$$

Ces deux vecteurs ont même moyenne ; examinons cependant leurs graphiques en bâtonnets, figurés à la *même échelle* sur l'axe vertical :

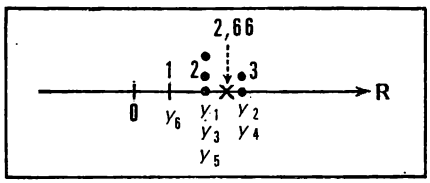


x a des écarts beaucoup plus grands que y autour de la moyenne ; un autre mode de figuration de x et de y montre

également que x est plus dispersé que y autour de sa moyenne :



x



y

Le calcul des variances (effectuez-le en vous rapportant chap. III, § 5-*c*) donne :

$$V_x = 104/6 = 17{,}33\ldots$$
$$V_y = 17/6 \simeq 2{,}8$$

La variance de y est beaucoup plus faible que celle de x.

La *variance* V_x d'un vecteur $\mathbf{x}$ de $\mathbf{R}^n$ est le carré divisé par n de la distance de $\mathbf{x}$ à sa moyenne ; V_x mesure la *dispersion moyenne* de $\mathbf{x}$ (autour de sa moyenne) :

$$V_x = \frac{(x_1 - m_x)^2 + \ldots + (x_n - m_x)^2}{n} = \frac{\| \mathbf{x} \|^2 - n.m_x^2}{n} = \frac{d^2(\mathbf{x}, \mathbf{m}_x)}{n}$$

Comme, dans le calcul de la variance, le carré de la distance d'un vecteur $\mathbf{x}$ à son vecteur moyen est divisé par n, dimension du vectoriel auquel appartient $\mathbf{x}$, ce calcul élimine l'influence que peut avoir, sur le résultat, ce nombre de coordonnées. Il serait absurde de comparer des longueurs, ou distances, de vecteurs n'ayant pas la même dimension.

Un vecteur de dimension 700, même très peu dispersé, peut avoir une longueur beaucoup plus grande qu'un vecteur de dimension 7 très dispersé, du seul fait que dans le premier cas il y a 700 nombres positifs à additionner, et 100 fois moins dans le second cas.

Les variances, qui *sont des moyennes*, qui divisent la somme des carrés des écarts par n, peuvent, elles, être comparées, même s'il s'agit de vecteurs qui n'ont pas même dimension.

Exemple :

Reprenons $\mathbf{x} = (-5, 2, 7, 4, 6, -1)$, $\mathbf{x} \in \mathbf{R}^6$.

Et soit : $\mathbf{z} = (-5, 5, 7, -1)$, $\mathbf{z} \in \mathbf{R}^4$.

On a :

$$d^2(\mathbf{x}, \mathbf{m}_x) = 104$$

Et :

$$d^2(\mathbf{z}, \mathbf{m}_z) = \mathbf{z}^2 - n.m_z = 100 - 9 = 91$$

La distance de $\mathbf{x}$ à son vecteur moyen est plus grande que celle de $\mathbf{z}$ à son vecteur moyen.

Mais :

$$V_x = 104/6 = 17{,}66...$$
$$V_z = 91/4 = 22{,}75$$

On conclut donc que $\mathbf{z}$, qui a une variance plus grande que $\mathbf{x}$, est plus dispersé que $\mathbf{x}$.

Mais c'est l'utilisation de la variance à la description d'un échantillon, et à l'*inférence statistique*, qui fournit la raison la plus profonde de cette division par n : il faut alors en effet une définition de la variance qui soit cohérente avec celle qui est adoptée en calcul des probabilités, et par conséquent qui soit pondérée avec des coefficients positifs de somme 1 (cf. § 2-*e*, et chap. IX, § 4).

Écart type :

La variance exprime la dispersion d'un tableau de nombres (d'un vecteur) par le carré d'une distance ; si l'on veut se ramener à une distance, on considère la racine carrée de la variance ; c'est ce qu'on appelle *l'écart type*, noté σ_x, du vecteur **x**.

Écart type d'un vecteur **x** : racine carrée de sa variance :

$$\sigma_x = (V_x)^{1/2}$$

*4. Métriques non linéaires

a) *D'autres façons de calculer la longueur d'un vecteur et des distances*

On a dit au § 1 qu'il y a d'autres manières de calculer la longueur d'un vecteur que de prendre la longueur euclidienne, qui a été considérée jusqu'ici.

Première manière :

La longueur d'un vecteur est la plus grande des *valeurs absolues* de ces coordonnées.

Exemple :

Le vecteur (1, — 3, 0, 2, — 1) a pour valeurs absolues de ses coordonnées 1, 3, 0, 2, 1 respectivement ; la plus grande est 3 ; sa longueur est 3.

D'une façon générale, si $\mathbf{x} = (x_1, x_2, \ldots, x_n) \in \mathbf{R}^n$, sa longueur $l_1(\mathbf{x})$ est :

$$l_1(\mathbf{x}) = \underset{1 \leqslant i \leqslant n}{\text{Max}} \mid x_i \mid$$

Ce qui se lit : la longueur de **x** est le *maximum*, lorsque l'indice i va de 1 à n, des valeurs absolues $\mid x_i \mid$ des coordonnées de **x**.

On prend encore pour distance la *longueur de la différence* $\mathbf{x} - \mathbf{y}$:

$$d_1(\mathbf{x}, \mathbf{y}) = l_1(\mathbf{x} - \mathbf{y}) = \max_{1 \leqslant i \leqslant n} \mid x_i - y_i \mid$$

Ceci signifie que la distance entre **x** et **y** est égale au *plus grand écart* entre coordonnées correspondantes de **x** et de **y**.

Seconde manière :

La longueur d'un vecteur est la somme de ses coordonnées prises en valeur absolue. Le vecteur $(1, -3, 0, 2, -1)$ a ainsi pour longueur :

$$1 + 3 + 0 + 2 + 1 = 7$$

Variante à cette manière :

Prendre la moyenne arithmétique des valeurs absolues des coordonnées.

D'une façon générale, on a cette fois-ci :

$$l_2(\mathbf{x}) = \mid x_1 \mid + \mid x_2 \mid + \ldots + \mid x_n \mid$$

$$\left(\text{ou } l'_2(\mathbf{x}) = \frac{\mid x_1 \mid + \ldots + \mid x_n \mid}{n}\right).$$

Et comme distance :

$$d_2(\mathbf{x}, \mathbf{y}) = l_2(\mathbf{x} - \mathbf{y}) = \mid x_1 - y_1 \mid + \ldots + \mid x_n - y_n \mid$$
$$d'_2(\mathbf{x}, \mathbf{y}) = l'_2(\mathbf{x}, \mathbf{y})$$

b) *Métriques*

Chacune des longueurs l_1 et l_2, et des distances d_1 et d_2 sont bien des longueurs et des distances en ce sens qu'elles satisfont l'une et l'autre aux propriétés essentielles de la longueur et des distances euclidiennes (§ 1-*d*), à savoir :

1) positivité ;
2 *bis*) symétrie ;
3) inégalité du triangle.

Une *distance* sur un ensemble E quelconque est une application $d : E \times E \to \mathbf{R}_0^+$, qui associe à tout couple (a, b) d'éléments de E un nombre $d(a, b)$ avec les propriétés :

1) $d(a, b) \geqslant 0$;
1 *bis*) $d(a, b) = 0 \Leftrightarrow a = b$;
2 *bis*) $d(b, a) = d(a, b)$;
3) $d(a, b) + d(b, c) \geqslant d(a, c)$.

On dit qu'on a défini une métrique sur E, ou que E est un *espace métrique*, lorsque E est muni d'une distance. Grâce à ces propriétés de la distance, celle-ci garde en effet l'essentiel des caractères qu'elle possède dans le cas où E est l'ensemble des points de la géométrie euclidienne, l'espace euclidien.

Exercice en guise d'exemple :

Prouver que sur l'ensemble $E = P(\Omega)$ des parties d'un ensemble *fini* Ω, la distance :

$$\delta(A, B) = |A \cap B'| + |A' \cap B|$$

(où $|X|$ désigne le nombre d'éléments d'une partie X de Ω, et X' la partie complémentaire de X dans Ω) satisfait bien aux propriétés 1, 1 *bis*, 2 *bis* et 3.

c) *Métriques sur des vectoriels*

Lorsque E est un *vectoriel*, ce qui est le cas dans ce chapitre, les métriques intéressantes sont d'abord celles qui établissent un lien entre la structure de vectoriel (addition des vecteurs ; produit d'un vecteur par un nombre) et la structure de **R** (addition des nombres ; produit des nombres ; comparaison des nombres).

Ce sont donc celles qui sont définies à partir d'une *norme* ou *longueur* des vecteurs qui satisfont à :

1) positivité : $l(\mathbf{x}) \geqslant 0$;

1 *bis*) $l(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = 0$;

2) lien entre les homothéties et le produit des nombres : $l(\lambda\mathbf{x}) = |\lambda| \, l(\mathbf{x})$;

2 *bis*) $l(-x) = l(\mathbf{x})$;

3) lien entre l'addition des *vecteurs*, celle des nombres, et la comparaison entre ceux-ci :

$$l(\mathbf{x} + \mathbf{y}) \leqslant l(\mathbf{x}) + l(\mathbf{y})$$

La distance est alors définie, dans *un vectoriel*, comme longueur de la différence :

$$l(\mathbf{x} - \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$$

Les distances d_1 et d_2, comme la distance euclidienne, sont toutes dans ce cas.

d) *La métrique euclidienne est linéaire*

— Mais il y a ensuite une prééminence de la distance euclidienne parce que, étant définie à partir d'un *produit scalaire*, c'est-à-dire d'une *forme bilinéaire* (chap. III, § 2), elle est la plus en accord avec la logique de l'algèbre et du calcul linéaire ; il conviendrait d'ailleurs plutôt de la qualifier de *distance linéaire* (et de parler de *métrique linéaire*) que de distance euclidienne.

e) *Valeur centrale à distance minimum pour les distances* d_1 *et* d_2

Les mots « meilleure approximation », « plus proche de », etc., n'ont de sens que par rapport à la manière de *mesurer la proximité*, c'est-à-dire *par rapport à la métrique adoptée.*

En particulier, la suite *constante* $\mathbf{t} = (t, t, \ldots, t)$, qui est à *distance minimum* d'une suite (vecteur) $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, et qui constitue par conséquent la meilleure valeur centrale de la suite des nombres $x_1, x_2, \ldots, x_n$, n'est la moyenne m_x de $(x_1, x_2, \ldots, x_n)$ que si la distance est la distance euclidienne.

Pour les distances d_1 et d_2, cette meilleure valeur centrale t est respectivement :

— *Pour* d_1 $\left(d_1(\mathbf{x}, \mathbf{y}) = \max\limits_{1 \leqslant i \leqslant n} |x_i - y_i|\right)$:

$$t = 1/2\left(\max_{1 \leqslant i \leqslant n} x_i + \min_{1 \leqslant i \leqslant n} x_i\right)$$

C'est la moyenne arithmétique entre la plus grande et la plus petite (en valeur *relative* et non absolue) des coordonnées de $\mathbf{x}$, le milieu de l'intervalle déterminé par les deux valeurs extrêmes prises par les x_i.

La mesure correspondante de la dispersion est l'*étendue* du vecteur $\mathbf{x} = (x_1, \ldots, x_n)$, c'est-à-dire la différence entre les deux valeurs extrêmes :

$$\max_{1 \leqslant i \leqslant n} x_i - \min_{1 \leqslant i \leqslant n} x_i$$

— *Pour* d_2 $(d_2(\mathbf{x}, \mathbf{y}) = |x_1 - y_1| + \ldots + |x_n - y_n|)$: t est alors la *médiane* de la suite $x_1, x_2, \ldots, x_n$ (*MSH*, chap. XII, § 5) si cette médiane existe ; n'importe quelle valeur de l'intervalle médian sinon (généralement, on choisit le milieu de celui-ci).

La mesure correspondante de la dispersion est :

— soit *l'intervalle interquartile médian :* celui qui a pour milieu la médiane et contient la moitié des n nombres, en général pas tous distincts, $x_1, x_2, \ldots, x_n$;

— soit plus généralement *l'intervalle interquantile médian :* celui qui a pour milieu la médiane, et contient la fraction $2/q$ des nombres $x_1, x_2, \ldots, x_n$.

N.B. — 1. La médiane peut être définie, comme valeur centrale, même si les x_i ne sont pas des nombres, mais seulement des éléments d'un ensemble totalement ordonné (cf. *MSH*, chap. XIII, § 5).

De la même façon, intervalle interquartile médian et intervalle interquantile médian sont des « mesures » (non numériques) de la dispersion lorsque les x_i ne sont pas des nombres, mais appartiennent à un ordre total : elles permettent de *comparer* les dispersions de plusieurs statistiques.

2. Si les distances d_1 et d_2 ont l'inconvénient d'être non linéaires, elles présentent par contre un grand avantage : leurs expressions étant rationnelles par rapport aux coordonnées des vecteurs (elles ne font pas intervenir, comme la distance euclidienne, de racine carrée), elles sont définies dans le corps $\mathbf{Q}$ des rationnels. De façon précise, pour le vectoriel $\mathbf{Q}^n$ des suites de n nombres rationnels, d_1 et d_2 sont toujours à valeurs dans $\mathbf{Q}$. Leur emploi ne nécessite donc pas de calcul numérique approché (cf. chap. I, § 8 et 9) lorsque les données sont des nombres rationnels.

3. En *inférence statistique* on utilise couramment, outre les distances *euclidiennes*, d_1 et d_2, une autre mesure de l'écart entre suites de nombres, connue sous le nom de *Khi-Deux*. Cet écart n'est pas une distance ; ce qui justifie son emploi, ce sont ses propriétés en calcul des probabilités.

Récapitulation

Longueur (norme) euclidienne d'un vecteur $\mathbf{x}$ de $\mathbf{R}^n$: racine carrée de la somme des carrés des coordonnées de $\mathbf{x}$.

Si :

$$\mathbf{x} = (x_1, x_2, \ldots, x^n)$$

sa longueur est :

$$|| \mathbf{x} || = (x_1^2 + x_2^2 + \ldots + x_n^2)^{1/2}$$

On a :

$$|| \mathbf{x} ||^2 = \mathbf{x}.\mathbf{x}$$

La distance est définie à partir du produit scalaire.

↓

Distance euclidienne de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ de $\mathbf{R}^n$: longueur de leur différence.

$$d(\mathbf{x},\mathbf{y}) = ||\mathbf{x}-\mathbf{y}|| = [(x_1-y_1)^2 + \ldots + (x_n-y_n)^2]^{1/2}$$

↓

Propriétés de la distance :

1. $d(\mathbf{x}, \mathbf{y}) \geqslant 0$ (la distance est positive) ;
1 *bis*. $d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$: la distance ne s'annule que si les vecteurs sont égaux ;
2 *bis*. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$ (symétrie) ;
3. $d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \geqslant d(\mathbf{x}, \mathbf{z})$: inégalité du triangle.

En statistique

Mesurer la *dispersion* d'une suite de nombres : distance moyenne du *vecteur* x correspondant à son vecteur moyen.

Ecart type : Distance de x à sa moyenne m_x, et changement d'unité $1/n^{1/2}$.

$$\sigma_x = \frac{d(x, m_x)}{n^{1/2}} = \frac{\| x - m_x \|}{n^{1/2}}$$

Variance : Carré de l'écart type :

$$V_x = \sigma_x^2 = \frac{d^2(x, m_x)}{n}$$

Calcul pratique de la variance et de l'écart type :

1. *Calculer la variance :*

$$V_x = \frac{\| x \|^2}{n} - m_x^2$$

$$\| x \|^2 = (x_1^2 + x_2^2 + \ldots + x_n^2)$$

$$m_x = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

2. Puis *calculer l'écart type* : $\sigma_x = V_x^{1/2}$.
Prendre la racine carrée de la variance.

CHAPITRE V

COSINUS ET CORRÉLATION LINÉAIRE

1. COSINUS DE DEUX VECTEURS

a) *Définition*

$\mathbf{x}$ et $\mathbf{y}$ étant deux vecteurs *non nuls* d'un *même vectoriel* V, on appelle cosinus de $\mathbf{x}$ et $\mathbf{y}$, noté cos $(\mathbf{x}, \mathbf{y})$, le rapport du produit scalaire de $\mathbf{x}$ et $\mathbf{y}$ au produit des longueurs de ces vecteurs :

Si $\mathbf{x} \neq \mathbf{0}$ et $\mathbf{y} \neq \mathbf{0}$:

$$\text{Cos}\,(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|}$$

Dans le vectoriel $\mathbf{R}^n$ de dimension n, l'expression de cos $(\mathbf{x}, \mathbf{y})$ en fonction des coordonnées est donc :

$$\text{Cos}\,(\mathbf{x}, \mathbf{y}) = \frac{x_1 y_1 + x_2 y_2 + \ldots + x_n y_n}{(x_1^2 + \ldots + x_n^2)^{1/2}(y_1^2 + y_2^2 + \ldots + y_n^2)^{1/2}}$$

Exemple :

Soit dans $\mathbf{R}^4$ les deux vecteurs :

$$\mathbf{x} = (-1, 3, -5, 1) \qquad \text{et} \qquad \mathbf{y} = (-1, 3, 4, 0)$$

on a :

$$\mathbf{x}.\mathbf{y} = -10 \qquad \|\mathbf{x}\| = 6 \qquad \|\mathbf{y}\| = (26)^{1/2} \simeq 5$$

$$\text{Cos}\,(\mathbf{x}, \mathbf{y}) = \frac{-10}{6.(26)^{1/2}} \simeq \frac{-10}{30} = -\frac{1}{3}$$

N.B. — Si l'un des deux vecteurs $\mathbf{x}$ ou $\mathbf{y}$ était nul, le numérateur et le dénominateur du rapport $\dfrac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|}$ seraient nuls ; ce rapport n'a alors plus de sens ; le cosinus n'est pas défini quand l'un des vecteurs est nul.

b) *La valeur du cosinus ne dépend pas des unités choisies : il est invariant dans les homothéties*

On se souvient qu'en ce qui concerne le produit scalaire multiplier deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ par un même nombre λ, c'est multiplier ce produit par λ^2 :

$$(\lambda\mathbf{x}.\lambda\mathbf{y}) = \lambda^2\,(\mathbf{x}.\mathbf{y}) \qquad \text{(chap. III, § 5)}$$

Il en résulte pour les longueurs et les distances (chap. IV, § 1) :

$$\|\lambda\mathbf{x}\| = |\lambda| \times \|\mathbf{x}\| \qquad d\,(\lambda\mathbf{x}, \lambda\mathbf{y}) = |\lambda|\, d\,(\mathbf{x}, \mathbf{y})$$

En description statistique, l'homothétie de rapport λ, la multiplication par un même nombre λ de toutes les coordonnées x_i de chaque vecteur $\mathbf{x}$, correspond en général à un *changement d'unité* ou d'*échelle* dans la mesure (le repérage) de celles-ci.

Ce qu'on vient de rappeler signifie donc que si l'on fait une homothétie de rapport λ sur un vectoriel, *toutes les distances sont multipliées dans le rapport* $|\lambda|$ (valeur absolue de λ).

Le cosinus, par contre, ne dépend pas des unités choisies ; il est *invariant par homothétie* ; en effet :

$$\text{Cos}\,(\lambda\mathbf{x}, \lambda\mathbf{y}) = \frac{(\lambda\mathbf{x}).(\lambda\mathbf{y})}{\|\lambda\mathbf{x}\| \times \|\lambda\mathbf{y}\|} = \frac{\lambda^2\,(\mathbf{x}.\mathbf{y})}{|\lambda|^2\,\|\mathbf{x}\| \times \|\mathbf{y}\|}$$

Comme : $\lambda^2 = |\lambda|^2$, on a finalement :

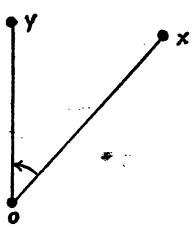
$$\text{Cos}\,(\lambda\mathbf{x}, \lambda\mathbf{y}) = \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|} = \text{Cos}\,(\mathbf{x}, \mathbf{y})$$

$$\text{Cos}\,(\lambda\mathbf{x}, \lambda\mathbf{y}) = \text{Cos}\,(\mathbf{x}, \mathbf{y})$$

Quand deux vecteurs sont multipliés par un même nombre, leur cosinus garde la même valeur : le cosinus ne dépend pas des unités choisies.

c) *Cosinus et angles ; cas remarquables*

— On démontre (et on admet ici sans démonstration) qu'en ce qui concerne les vecteurs de $\mathbf{R}^2$ ou $\mathbf{R}^3$, dont on a vu qu'ils peuvent se représenter respectivement par des points du plan ($\mathbf{R}^2$) ou de l'espace ($\mathbf{R}^3$), c'est bien le cosinus trigonométrique de l'*angle* des deux vecteurs géométriques $\mathbf{x}$ et $\mathbf{y}$ qui se calcule par la règle :



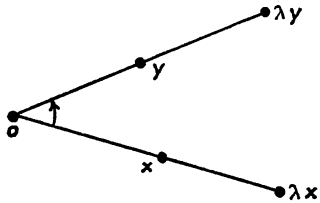
$$\text{Cos}\,(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|}$$

$$\left(= \frac{x_1 y_1 + x_2 y_2}{(x_1^2 + x_2^2)^{1/2}\,(y_1^2 + y_2^2)^{1/2}} \text{ dans } \mathbf{R}^2\right)$$

Cet angle ne *dépend évidemment pas des unités choisies dans la mesure des coordonnées* ; si l'on multiplie les vecteurs $\mathbf{x}$

et y par λ, comme λx est *aligné avec* x et λy aligné avec y, l'angle de λx et λy est le même que l'angle de x et de y :

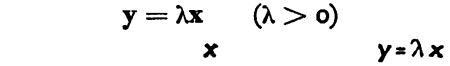
$$\text{Cos}\,(x, y) = \text{Cos}\,(\lambda x, \lambda y)$$



— Le cosinus défini § 1-*a* possède une autre propriété du cosinus trigonométrique d'un angle en géométrie euclidienne ; on sait, en effet, que celui-ci est toujours compris entre — 1 et + 1, et qu'il vaut :

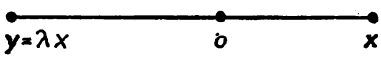
— 1 si les deux vecteurs sont alignés et de même sens :

$$y = \lambda x \qquad (\lambda > 0)$$

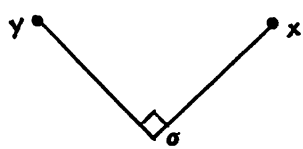


— — 1 si les deux vecteurs sont alignés et de sens contraire :

$$y = \lambda x \qquad (\lambda < 0)$$



— 0 si les deux vecteurs sont perpendiculaires (orthogonaux).



Toutes ces propriétés restent vraies pour des vecteurs d'un même vectoriel de dimension quelconque, et le calcul du cosinus permet donc d'y généraliser le concept d'angle de deux vecteurs, comme la définition adoptée pour la distance a permis de généraliser celle-ci (chap. III et IV).

— En effet, dans le cas général de vecteurs de dimension n, de vecteurs $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ ayant n coordonnées (vecteurs de $\mathbf{R}^n$), le cosinus est *toujours compris entre 1 et — 1*, et *vaut 1 ou — 1 lorsque l'un des vecteurs est le produit de l'autre par un nombre.*

Ceci résulte de l'inégalité de Schwarz :

$$-\|\mathbf{x}\| \times \|\mathbf{y}\| \leqslant \mathbf{x}.\mathbf{y} \leqslant \|\mathbf{x}\| \times \|\mathbf{y}\|$$

Elle est valable quelle que soit la dimension du vectoriel, et pour tous les couples x et y de vecteurs d'un même vectoriel. Si l'on divise chaque terme par $\|\mathbf{x}\| \times \|\mathbf{y}\|$ l'inégalité de Schwarz s'écrit encore :

$$-1 \leqslant \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|} \leqslant +1$$

c'est-à-dire : $\boxed{-1 \leqslant \cos(\mathbf{x}, \mathbf{y}) \leqslant +1}$

L'inégalité de Schwarz signifie que *le cosinus de deux vecteurs non nuls est compris entre — 1 et 1.*

— Si deux vecteurs x et y sont *orthogonaux*, ceci signifie par définition (chap. III, § 4) que :

$$\mathbf{x}.\mathbf{y} = 0$$

Donc que : $\cos(\mathbf{x}, \mathbf{y}) = \dfrac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|} = 0.$

— Si le vecteur $\mathbf{y}$ est *homothétique* du vecteur $\mathbf{x}$ dans le rapport λ, on a :

$$\mathbf{y} = \lambda\mathbf{x}$$

Donc :

$$(\mathbf{x}.\mathbf{y}) = \lambda\,(\mathbf{x}.\mathbf{x}) = \lambda \,||\, \mathbf{x} \,||^2$$
$$||\, \mathbf{y} \,|| = ||\, \lambda\mathbf{x} \,|| = |\, \lambda \,|\, ||\, \mathbf{x} \,||$$

D'où :

$$\cos\,(\mathbf{x}, \mathbf{y}) = \frac{\lambda \,||\, \mathbf{x} \,||^2}{|\, \lambda \,|\, ||\, \mathbf{x} \,||^2} = \frac{\lambda}{|\, \lambda \,|}$$

Si $\lambda > 0$: $|\, \lambda \,| = \lambda$ et $\dfrac{\lambda}{|\, \lambda \,|} = 1$

Si $\lambda < 0$: $|\, \lambda \,| = -\lambda$ et $\dfrac{\lambda}{|\, \lambda \,|} = -1$

On a donc bien le résultat :

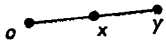
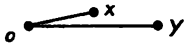
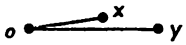
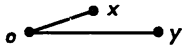
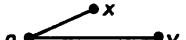
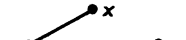
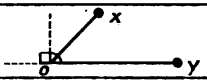
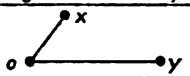
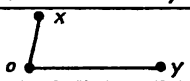
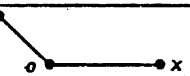
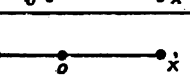
> Lorsque $\mathbf{y}$ est le produit du vecteur $\mathbf{x}$ par un nombre λ :
>
> $\mathrm{Cos}\,(\mathbf{x}, \mathbf{y}) = 1$ si $\lambda > 0$
> $\mathrm{Cos}\,(\mathbf{x}, \mathbf{y}) = -1$ si $\lambda < 0$

On démontre que réciproquement :

— si $\cos\,(\mathbf{x}, \mathbf{y}) = 1$, alors $\mathbf{y} = \lambda\mathbf{x}$, et $\lambda > 0$;
— si $\cos\,(\mathbf{x}, \mathbf{y}) = -1$, alors $\mathbf{y} = \lambda\mathbf{x}$, et $\lambda < 0$.

On peut donc dire :

> Le cosinus de deux vecteurs d'un même vectoriel est égal à 1 ou — 1 *si et seulement si* l'un de ces vecteurs est le produit de l'autre par un nombre.

Cosinus	Angle	Figure géométrique
1	0°	
0,9990...	2°	
0,990...	8°	
0,950...	18°	
0,90...	25°	
$\frac{\sqrt{3}}{2}$ = 0,866...	30°	
$\frac{\sqrt{2}}{2}$ = 0,70 ...	45°	
$\frac{1}{2}$ = 0,50	60°	
0,17...	80°	
0	90°	
$-\frac{1}{2}$	120°	
- 1	180°	

d) *Table de cosinus*

Des tables donnent la correspondance entre valeur du cosinus et valeur de l'angle de deux vecteurs de la géométrie. Lorsque le cosinus est presque égal à 1, l'angle est *très petit* ; si le cosinus est proche de — 1, l'angle est *presque plat* ; si le cosinus est voisin de 0, l'angle est *presque droit*. Une table réduite précise (p. 144) cette correspondance ; il est recommandé de s'y reporter, et de faire la traduction du cosinus à l'angle, chaque fois qu'on calcule un coefficient de corrélation linéaire (§ 3) qui, on le verra, est un cosinus.

2. COSINUS ET AJUSTEMENT LINÉAIRE

a) *Ajustements et ajustement linéaire*

En description statistique, les vecteurs que l'on traite sont toujours des suites de nombres (obtenues par des mesures ou des dénombrements), donc des applications d'un ensemble $\mathrm{E} = \{ 1, 2, \ldots, n \}$ dans $\mathbf{R}$ (cf. *MSH*, chap. XI).

$$\mathbf{x} = (x_1, x_2, \ldots, x_n) \qquad \mathbf{x} : \mathrm{E} \to \mathbf{R}$$

x_1 est l'image de 1, x_2 celle de 2, etc.

Lorsque l'on a deux vecteurs, x et y, d'un *même vectoriel* de dimension n, résultant par exemple de deux séries de mesures sur les éléments de $\mathrm{E} = \{ 1, 2, \ldots, n \}$, on recherche souvent s'il y a une *liaison* entre les x_i et les y_i ; par exemple si les y_i sont, au moins approximativement, fonction des x_i :

$$y_i = f(x_i) \qquad i \in \{ 1, 2, \ldots, n \}$$

On parle alors d'*ajustement* de la liaison par la fonction f.

Parmi les fonctions possibles, il y a la *fonction linéaire* : y_i varie linéairement en fonction de x_i, c'est-à-dire propor-

tionnellement à x_i ; si le coefficient de proportionnalité est λ (c'est le coefficient a dans les notations du chap. II), cela signifie que :

$$y_i = \lambda x_i \qquad (i \in \{1, 2, \ldots, n\})$$

c'est-à-dire que le vecteur **y** est le produit du vecteur **x** par le nombre λ (**y** est homothétique à **x** dans le rapport λ).

Si **y** est le produit de **x** par le nombre λ, on a :

$$y_1 = \lambda x_1, \qquad y_2 = \lambda x_2, \qquad \ldots, \qquad y_n = \lambda x_n$$

y_i varie proportionnellement à x_i, y_i est *fonction linéaire* de x_i.

On a vu au chapitre II, § 5 et 6, que d'autres types de liaison : y_i fonction exponentielle de x_i, ou y_i fonction logarithme de x_i, ou y_i fonction puissance de x_i ($y_i = a^{x_i}$, $y_i = \log_a x_i$, $y_i = x_i^a$) pouvaient se ramener à des fonctions *linéaires* ou *affines* par des changements d'échelle ; d'où l'importance fondamentale d'avoir un instrument qui permette de *mesurer le degré de liaison linéaire* entre deux « variables » numériques, entre deux vecteurs d'un même vectoriel.

Cet instrument, c'est le cosinus.

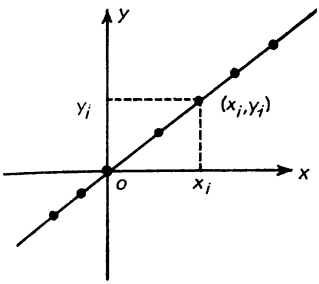
Pour la fonction affine, dont on a souvent besoin, c'est encore un cosinus qui sert ; on l'étudiera au paragraphe suivant.

b) *Cas d'une fonction linéaire*

On peut, comme au chapitre II, figurer les deux vecteurs **x** et **y** sur un *graphique cartésien :* un axe pour le vecteur **x**, le second pour **y**, et chaque couple (x_i, y_i) est représenté par un point.

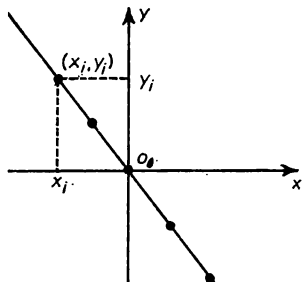
On a vu (chap. II, § 6) que si y_i est fonction linéaire de x_i, c'est-à-dire si $y = \lambda x$, les points du graphique sont alignés.

Or, en ce cas, le cosinus de x et de y vaut 1 ou -1 selon que λ, la pente de la droite, le coefficient de la fonction linéaire liant x_i à y_i, est positif ou négatif.



$\text{Cos}(\mathbf{x}, \mathbf{y}) = 1$

$y_i = \lambda x_i, \lambda > 0$



$\text{Cos}(\mathbf{x}, \mathbf{y}) = -1$

$y_i = \lambda x_i, \lambda < 0$

y_i est *fonction linéaire* de x_i si et seulement si cos (x, y) vaut 1 ou -1.

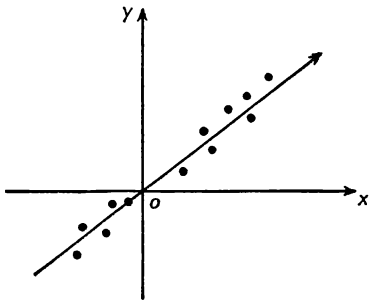
c) *Cas d'une liaison approximativement linéaire (droite d'ajustement)*

Il est exceptionnel que l'on obtienne dans deux séries de mesures une liaison absolument linéaire, c'est-à-dire que les points du graphique soient strictement alignés ; par contre il est fréquent que les points soient approximativement alignés, que la liaison entre x_i et y_i soit presque linéaire ; les points du graphique formeront alors un « nuage » allongé autour d'une droite. Dans ce cas, cos (x, y) est presque égal à 1 si la droite qui constitue une approximation de la

liaison entre x_i et y_i est de pente positive ($\lambda > 0$), cos (x, y) est presque égal à — 1 si cette pente est négative.

Cette droite est appelée *droite d'ajustement.*

Plus précisément, on peut chercher pour quelle valeur du coefficient a la *distance* entre le vecteur y et un vecteur ax, proportionnel à x dans le rapport a, est minimum.



Cos (x, y) proche de 1

$y_i \simeq \lambda x_i$

C'est-à-dire encore, quelle est la fonction linéaire $Y = aX$, la plus proche du nuage des points (x_i, y_i).

Il faut, pour cela, choisir a tel que $|| \mathbf{y} - a\mathbf{x} ||^2$ soit minimum. Un calcul bien souvent fait (chap. III), donne :

$$|| \mathbf{y} - a\mathbf{x} ||^2 = || \mathbf{y} ||^2 - 2\, a\, (\mathbf{x}.\mathbf{y}) + a^2 || \mathbf{x} ||^2$$

Ce trinôme, du second degré en a, est à son minimum pour :

$$a = \frac{\mathbf{x}.\mathbf{y}}{|| \mathbf{x} ||^2}$$

C'est-à-dire :

$$\boxed{a = \frac{|| \mathbf{x} ||}{|| \mathbf{y} ||} \cos (\mathbf{x}, \mathbf{y})}$$

On voit par cette expression en quel sens le cosinus est une mesure du *degré de linéarité* dans la liaison entre deux variables, deux vecteurs $\mathbf{x}$ et $\mathbf{y}$:

> Si $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ et $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, la fonction linéaire $Y = aX$ la plus proche (pour la métrique euclidienne) de l'ensemble des couples (x_i, y_i) est celle qui a pour coefficient :
>
> $$a = \frac{\| \mathbf{x} \|}{\| \mathbf{y} \|} \cos (\mathbf{x}, \mathbf{y})$$

Exemple :

$\mathbf{x} = (1, 0, 3, -2)$ $\qquad$ $\mathbf{y} = (0{,}4 ; -0{,}1 ; 1{,}6 ; -0{,}9)$

(on a séparé les coordonnées de y par « ; » pour ne pas confondre avec la virgule de l'écriture décimale des nombres).

y semble approximativement égal à $1/2\,\mathbf{x}$:

$$1/2\,\mathbf{x} = (0{,}5 ; 0 ; 1{,}5 ; -1)$$

On trouve : $\mathbf{x}.\mathbf{y} = 6$ $\quad \| \mathbf{x} \|^2 = 14$; $\| \mathbf{y} \|^2 = 3{,}4$.

D'où :

$$\cos (\mathbf{x}, \mathbf{y}) \simeq 0{,}85 \quad \text{et} \quad a = \frac{\| \mathbf{x} \|}{\| \mathbf{y} \|} \cos (\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}.\mathbf{y}}{\| \mathbf{x}^2 \|} = \frac{3}{7}$$

0,85 correspond à un angle de 30° environ : l'approximation de la liaison entre x_i et y_i par une fonction linéaire est *mauvaise.* La meilleure droite d'ajustement a pour équation : $Y = 3/7\,X$.

> Lorsque $\cos (\mathbf{x}, \mathbf{y})$ est proche de 1 (ou de -1), y_i est *approximativement* fonction *linéaire* de x_i ; l'approximation est d'autant meilleure que $\cos (\mathbf{x}, \mathbf{y})$ est proche de 1 (ou de -1).
>
> L'angle des vecteurs $\mathbf{x}$ et $\mathbf{y}$ est alors *petit* (ou presque *plat*).

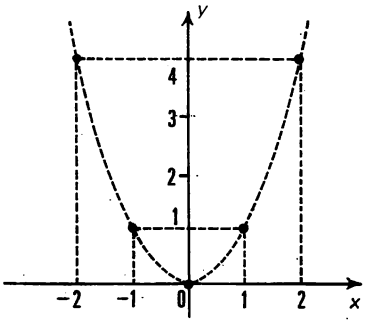
d) *Cas où la liaison n'est pas linéaire*

Lorsque cos (x, y) est *éloigné de 1 ou de — 1*, en particulier lorsqu'il est *voisin de* 0, ceci signifie que la liaison entre y_i et x_i, s'il y en a une, n'est *certainement pas linéaire ;* c'est le cas où l'angle des deux vecteurs x et y est presque droit. Mais *cela ne signifie pas nécessairement l'absence de liaison.*

Exemples :

a) $\mathbf{x} = (2, -1, 0, 1, -2)$;

$\mathbf{y} = (4, 1, 0, 1, 4)$.



On voit que :

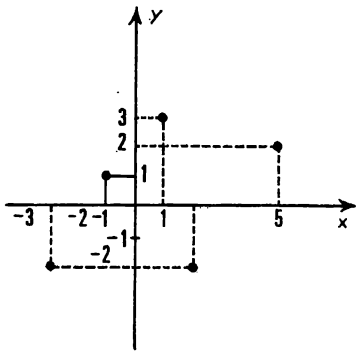
$$y_i = x_i^2$$

y_i est *fonction puissance* de x_i ; les points du graphique sont sur une parabole.

On a : $\mathbf{x}.\mathbf{y} = 8 - 1 + 0 + 1 - 8 = 0$.

Donc : $\mathbf{x} \perp \mathbf{y}$ et $\cos(\mathbf{x}, \mathbf{y}) = 0$.

b) $\mathbf{x} = (1, 3, -1, 2, 5, -3)$;
$\mathbf{y} = (3, 1, -2, 2, -2)$.



Le graphique ne fait pas apparaître de liaison entre x et y ; en tout cas, les points ne sont pas alignés, même approximativement.

On a : $\mathbf{x}.\mathbf{y} = 14$.

$$\| \mathbf{x} \| = 2\sqrt{10} \qquad \| \mathbf{y} \| = \sqrt{22}$$

Donc : $\cos(\mathbf{x}, \mathbf{y}) \simeq 0{,}4$, ce qui correspond d'ailleurs à un angle presque égal à 70° pour les vecteurs x et y.

On peut donc retenir :

$\cos(\mathbf{x}, \mathbf{y})$ est une mesure du *degré de linéarité* dans la liaison de y_i et de x_i ;

$\cos(\mathbf{x}, \mathbf{y}) = 1$ (ou -1) $\Leftrightarrow$ y_i est *fonction linéaire* de x_i ;

$\cos(\mathbf{x}, \mathbf{y}) \simeq 1$ (ou -1) $\Leftrightarrow$ y_i est *approximativement fonction linéaire* de x_i ;

$\cos(\mathbf{x}, \mathbf{y}) \simeq 0 \Rightarrow$ y_i *n'est pas fonction linéaire* de x_i.

3. Coefficient de corrélation linéaire et fonctions affines

a) *Définition*

Si l'on cherche à ajuster la liaison entre deux suites $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ et $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ non plus selon une *fonction linéaire*, mais selon une *fonction affine* — $Y = aX + b$ — (chap. II, § 1 *bis*), une mesure de la valeur d'une telle approximation est donnée par le coefficient de corrélation linéaire.

On appelle *coefficient de corrélation linéaire* (ou encore de *Bravais et Pearson*) de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$, le cosinus des vecteurs $\mathbf{x} - \mathbf{m}_x$ et $\mathbf{y} - \mathbf{m}_y$, où $\mathbf{m}_x$ est le *vecteur moyen* de $\mathbf{x}$ (toutes ses coordonnées sont égales à la moyenne arithmétique de celles de $\mathbf{x}$), $\mathbf{m}_y$ celui de $\mathbf{y}$. Le coefficient de corrélation linéaire se note r_{xy}.

$$r_{xy} = \cos(\mathbf{x} - \mathbf{m}_x, \mathbf{y} - \mathbf{m}_y)$$

Le coefficient de corrélation linéaire s'exprime au moyen de la covariance et des écarts types :

$$\cos(\mathbf{x} - \mathbf{m}_x, \mathbf{y} - \mathbf{m}_y) = \frac{(\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y)}{\| \mathbf{x} - \mathbf{m}_x \| \, \| \mathbf{y} - \mathbf{m}_y \|}$$

$$= \frac{1/n \, (\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y)}{1/n^{1/2} \, \| \mathbf{x} - \mathbf{m}_x \| \times 1/n^{1/2} \, \| \mathbf{y} - \mathbf{m}_y \|} = \frac{\mathrm{Cov}_{xy}}{\sigma_x \, \sigma_y}$$

On peut donc *calculer* le coefficient de corrélation linéaire r_{xy} comme rapport de la covariance de $\mathbf{x}$ et de $\mathbf{y}$ au produit de leurs écarts types :

$$r_{xy} = \frac{\mathrm{Cov}_{xy}}{\sigma_x \, \sigma_y} \qquad \left(= \frac{\mathrm{Cov}_{xy}}{V_x^{1/2} \, V_y^{1/2}} \right)$$

Ce dernier mode de calcul n'est économique que si, pour une raison quelconque, on a déjà dû calculer la covariance et les écarts types ; sinon, on fait le calcul direct.

b) *Calcul pratique du coefficient de corrélation linéaire*

On a vu au chapitre III, § 5, que :

$$(\mathbf{x} - \mathbf{m}_x).(\mathbf{y} - \mathbf{m}_y) = \mathbf{x}.\mathbf{y} - nm_x m_y$$

où : n est la dimension du vectoriel (le nombre des coordonnées) ;

$m_x = \dfrac{x_1 + \ldots + x_n}{n}$ est la moyenne arithmétique de $\mathbf{x}$;

$m_y = \dfrac{y_1 + y_2 + \ldots + y_n}{n}$ celle de $\mathbf{y}$.

De même :

$$\sigma_x^2 = V_x = \| \mathbf{x} - \mathbf{m}_x \|^2 = \| \mathbf{x} \|^2 - nm_x^2$$
$$\sigma_y^2 = V_y = \| \mathbf{y} \|^2 - nm_y^2$$

On a donc :

$$r_{xy} = \frac{\mathbf{x}.\mathbf{y} - nm_x m_y}{(\| \mathbf{x} \|^2 - nm_x^2)^{1/2} (\| \mathbf{y} \|^2 - nm_y^2)^{1/2}}$$

Voici, page suivante, une disposition des calculs, pour deux vecteurs de $\mathbf{R}^5$.

i	1	2	3	4	5	Sommes		
x_i ...	3	— 1	2	0	1	$5 (= nm_x)$	$m_x = 5/5 = 1$	$nm_x^2 = 5 \times 1 = 5$
y_i ...	— 2	1	0	2	2	$3 (= nm_y)$	$m_y = 3/5$	$nm_y^2 = 5 \times 9/5^2 = 9/5$
$x_i y_i$ ..	— 6	— 1	0	0	2	$-5 = \mathbf{x}.\mathbf{y}$	$m_x m_y = 1 \times 3/5 = 3/5$	$nm_x m_y = 5 \times 3/5 = 3$
x_i^2 ...	9	1	4	0	1	$15 = \Vert \mathbf{x} \Vert^2$	$\Vert \mathbf{x} \Vert^2 - nm_x^2 = 15 - 5 = 10$	
y_i^2 ...	4	1	0	4	4	$13 = \Vert \mathbf{y} \Vert^2$	$\Vert \mathbf{y} \Vert^2 - nm_y^2 = 13 - 9/5 = 56/5$	

$$\mathbf{x}.\mathbf{y} - nm_x m_y = -5 - 3 = -8$$

$$r_{xy} = \frac{-8}{10^{1/2}\left(\frac{56}{5}\right)^{1/2}} = \frac{-8}{2^{1/2}.56^{1/2}} = \frac{-8}{4.7^{1/2}} = \frac{-2}{7^{1/2}} = \frac{-2}{7^{1/2}} \simeq \frac{-2}{2{,}6}$$

$$r_{xy} \simeq -0{,}7$$

Signification géométrique de ce résultat : — 0,7 est le cosinus de l'angle de 135° ; les vecteurs $\mathbf{x} - \mathbf{m}_x$ et $\mathbf{y} - \mathbf{m}_y$ font un angle voisin de 135° : on est très loin d'une liaison linéaire entre $(\mathbf{x} - \mathbf{m}_x)$ et $(\mathbf{y} - \mathbf{m}_y)$.

N.B. — Ce calcul peut également se faire en utilisant les expressions de la covariance et des variances au moyen des sommes S_x $(= nm_x)$ et S_y $(= nm_y)$ vues au § 5 du chapitre III, à la place de m_x et m_y respectivement.

c) *Ajustement par une fonction affine*

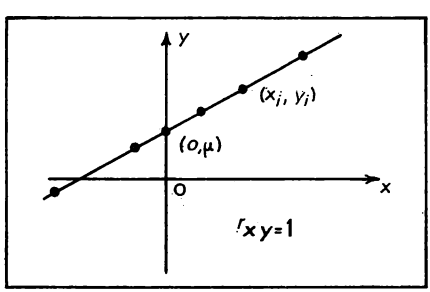
Comme l'a montré la discussion du § 2, ce que mesure r_{xy}, c'est le degré de liaison linéaire entre les « variables » $(x_i - m_x)$ et $(y_i - m_y)$.

$$r_{xy} = 1 \quad \Leftrightarrow (y_i - m_y) = \lambda\,(x_i - m_y), \qquad \lambda > 0$$
$$r_{xy} = -1 \Leftrightarrow (y_i - m_y) = \lambda\,(x_i - m_x), \qquad \lambda < 0$$

La relation : $y_i - m_y = \lambda\,(x_i - m_x)$
s'écrit encore : $y_i = \lambda x_i + (m_y - \lambda m_x)$
On a donc, en posant : $m_y - \lambda m_x = \mu$:

$$r_{xy} = 1 \text{ (ou } -1) \Leftrightarrow y_i = \lambda x_i + \mu$$

y_i est alors *fonction affine* de x_i ; graphiquement, les points (x_i, y_i) sont alignés sur une droite, qui ne passe plus nécessairement par l'origine, mais qui passe par le *point moyen* de coordonnées m_x et m_y dans le système des deux axes x et y. Elle coupe l' « axe des y » au point $(0, \mu)$ d'ordonnée μ (chap. II, § 1 *bis* et 6).



λ : pente de la droite
μ : ordonnée à l'origine

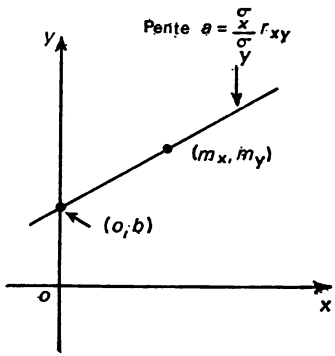
$$y_i = \lambda x_i + \mu \qquad \mu = m_y - \lambda m_x$$

Si r_{xy} est proche de 1 (ou de — 1) les points du graphique sont approximativement alignés, on a approximativement, pour y_i en fonction de x_i :

$$y_i \simeq \lambda x_i + \mu$$

La liaison entre les x_i et les y_i peut s'ajuster par une fonction affine, c'est-à-dire du type :

$$Y = aX + b \qquad (a = \lambda, b = \mu)$$



On montre, comme en 2 ci-dessus, que le *meilleur ajustement* est celui pour lequel le coefficient a, la pente de la droite d'ajustement, est :

$$\boxed{a = \frac{\sigma_x}{\sigma_y} r_{xy}}$$

σ_x et σ_y sont en effet proportionnels aux *longueurs* des vecteurs $\| \mathbf{x} - \mathbf{m}_x \|$ et $\| \mathbf{y} - \mathbf{m}_y \|$; au § 2, c'est le rapport des longueurs qui était le coefficient du cosinus dans l'expression de a.

On remarque d'ailleurs que la droite d'ajustement passe par le point de coordonnées (m_x, m_y), c'est-à-dire par le *point moyen* du nuage des points (x_i, y_i) ; la valeur b de l'ordonnée à l'origine de la droite d'ajustement $(Y = aX + b)$ est encore donnée par :

$$b = m_y - am_x$$

Si $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ et $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, le meilleur ajustement pour la métrique euclidienne par une fonction *affine* $(Y = aX + b)$ de la liaison entre x_i et y_i est obtenu en prenant :

$$a = \frac{\sigma_x}{\sigma_y} r_{xy} \qquad b = m_y - am_x$$

Droites de régression :

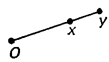
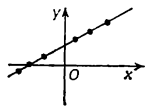
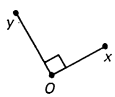
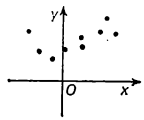
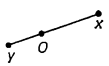
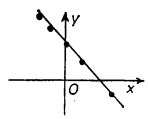
La droite d'équation $Y = aX + b$ qui constitue le meilleur ajustement est souvent appelée *droite de régression de Y en X ;* elle a été obtenue en considérant les y_i comme fonction (approximativement affine) des x_i, considérés comme « variables ».

Si l'on considère les y_i comme la variable, on obtient un ajustement par une droite d'équation $X = a' Y + b'$, dite de *régression de X en Y : $a' = (\sigma_y/\sigma_x) r_{xy}$, $b' = m_x - a' m_y$.*

Ces deux droites sont *distinctes,* et d'autant plus proches entre elles que r_{xy} est proche de 1 (ou — 1) ; toutes deux passent par le point moyen (m_x, m_y).

— Si r_{xy} est éloigné de 1 et — 1, en particulier si r_{xy} est proche de zéro, le graphique ne peut être ajusté selon une droite : cela *ne signifie pas toujours l'absence de liaison* entre y_i et x_i ; il peut y avoir une liaison, mais elle n'est pas affine (cf. § 2-*d*).

En résumé :

r_{xy}	Angle	Vecteurs	Type de la liaison	Graphique
1	0°		$y_i = \lambda x_i + \mu$ $\lambda > 0$ fonction *affine* (linéaire si $\mu = 0$)	
0	90°		Pas de liaison *linéaire* ou *affine*	
— 1	180°		$y_i = \lambda x_i + \mu$ $\lambda < 0$	

Ce que mesure r_{xy} :

Le degré de liaison *affine* entre **x** et **y** ;
ou encore :
Le degré de liaison *linéaire* entre $(\mathbf{x} - \mathbf{m}_x)$ et $(\mathbf{y} - \mathbf{m}_y)$.

N.B. — Le coefficient r_{xy} de Bravais et Pearson est qualifié de *linéaire* parce qu'il mesure, comme on vient de le voir, le degré de liaison linéaire ou affine entre **x** et **y** ; mais aussi, et surtout, parce qu'il est lié essentiellement, de par

sa définition au moyen du produit scalaire, à l'*algèbre linéaire.*

De même qu'il y a des métriques non linéaires (chap. IV, § 4), on peut définir pour certaines questions de statistique des coefficients de corrélation non linéaires. C'est ainsi par exemple que lorsque les variables en jeu ne sont pas des nombres, mais des rangs (appartiennent à un ensemble totalement ordonné), on utilise un coefficient tel que le coefficient τ_{xy} *(tau)* de Kendall, qui ne sera pas défini ici.

Récapitulation

Cosinus de deux vecteurs $\mathbf{x}$ et $\mathbf{y}$ *non nuls* d'un même vectoriel :

$$\cos(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{x}.\mathbf{y}}{\|\mathbf{x}\| \times \|\mathbf{y}\|}$$

$$= \frac{x_1 y_1 + x_2 y_2 + \ldots + x_n y_n}{(x_1^2 + y_2^2 + \ldots + x_n^2)^{1/2} (y_1^2 + \ldots + y_n^2)^{1/2}}$$

En statistique

Coefficient de corrélation linéaire :

$$r_{xy} = \cos(\mathbf{x} - \mathbf{m}_x, \mathbf{y} - \mathbf{m}_y) = \frac{\text{Cov } xy}{\sigma_x \sigma_y}$$

r_{xy} mesure le degré de liaison affine ($Y = aX + b$) entre les x_i et les y_i.

$r_{xy} = 1$ (ou -1) $\Leftrightarrow$ y_i fonction affine de x_i ;

$r_{xy} \simeq 1$ (ou -1) $\Leftrightarrow$ y_i approximativement fonction affine de x_i ;

$r_{xy} \simeq 0$ pas de liaison affine.

La meilleure droite d'ajustement *(droite de régression)* :

— de Y en X :

$$Y = aX + b$$

$$a = \frac{\sigma_x}{\sigma_y} r_{xy} \qquad b = m_y - a m_x$$

— de X en Y :

$$X = a' Y + b'$$

$$a' = \frac{\sigma_y}{\sigma_x} r_{xy} \qquad b' = m_x - a' m_y$$

TROISIÈME PARTIE

LA MESURE EN PROBABILITÉS

CHAPITRE VI

MESURES DE PROBABILITÉ

1. CALCUL DES POSSIBILITÉS (OU ALGÈBRE DES ÉVÉNEMENTS)

a) *Possibilités et probabilités*

L'expérimentateur se trouve souvent dans la situation suivante : il peut prévoir quels sont les résultats *possibles* de son expérience, mais non quel est, parmi ces possibles, celui qui se réalisera ; plus précisément, il peut déterminer *l'ensemble* des résultats possibles, l'ensemble des *possibilités*. Mais celle de ces possibilités qui se réalisera effectivement lui est inconnue avant que l'expérience soit faite.

Si par exemple on procède à une expérience sur des temps de réaction à un stimulus donné, on peut prévoir quels sont les temps de réactions possibles ; mais on ne connaît pas à l'avance ceux qui seront effectivement fournis par les sujets soumis à l'expérience.

De même, si l'on fait une enquête par sondage au sein d'une population donnée, les règles selon lesquelles l'échantillon que l'on observera est extrait de la population et la connaissance du questionnaire (s'il s'agit d'une enquête par questionnaire) permettent en général de construire l'ensemble des résultats possibles du sondage ; on ne peut dire à l'avance lequel de ces résultats sera en fait observé.

Le calcul des probabilités fournit à l'expérimentateur, dans des conditions précises, l'outil qui permet de juger du degré d'adéquation entre l'observation faite, et les résultats qu'il attendait de son expérience avant que celle-ci soit effectuée. La mise en œuvre de cet outil, c'est l'*inférence statistique*. Ce calcul s'exerce donc sur *l'ensemble des résultats possibles* de l'expérience. Et préalablement à ce calcul, qui est « quantitatif », puisqu'il attache à chaque résultat possible une *mesure numérique* du « degré de vraisemblance » de ce résultat, il y a un calcul « qualitatif », qui porte uniquement sur les résultats possibles eux-mêmes, ce que l'on appelle ici calcul des possibilités ou algèbre des événements. En quoi ce calcul consiste-t-il ?

b) *Ensemble des possibilités*

La première chose à faire est de déterminer sans ambiguïté *l'ensemble des possibilités ;* cet ensemble est désigné dans la suite par la lettre Ω (grand oméga).

Exemples :

1) De la population constituée par les étudiants de première année en Sociologie à la Faculté des Lettres de Paris, on extrait un *échantillon* (c'est-à-dire une *partie* de cette population) de 100 étudiants ; on s'intéresse à la proportion, dans l'échantillon, des étudiants qui ont fait au moins 6 ans d'anglais pendant leurs études secondaires.

L'ensemble Ω des possibilités est ici constitué par 101 valeurs possibles de cette proportion :

$$\Omega = \{ 0\ \%, 1\ \%, 2\ \%, \ldots, 99\ \%, 100\ \% \}$$

2) D'une urne contenant des boules noires et des boules blanches on extrait un échantillon de 12 boules ; on s'intéresse au nombre des boules blanches dans l'échantillon.

L'ensemble Ω est constitué par les treize valeurs possibles de ce nombre :

$$\Omega = \{ 0, 1, 2, \ldots, 11, 12 \}$$

3) De la même urne, on extrait une à une 12 boules, et on s'intéresse à l'ordre dans lequel apparaissent les blanches et les noires. Cette fois-ci Ω est constitué par les 2^{12} ($= 4\,096$) suites :

BB...BB, BB...BN, BB...BNB, ..., NNN...N

On peut encore dire que l'on obtient Ω en écrivant, dans l'ordre alphabétique par exemple, les 2^{12} mots de 12 lettres possibles avec un alphabet de 2 lettres, B et N.

Exercice :

Prouvez que le nombre de ces mots est bien 2^{12} ; pourquoi leur ensemble fournit-il l'ensemble Ω cherché ?

Il peut arriver que l'ensemble des possibilités soit non *fini*, comme dans les exemples précédents, mais infini ; dans ce cours, sauf mention explicite du contraire, on supposera toujours que *l'ensemble* Ω *est fini :* il n'y a qu'un nombre fini de possibilités.

c) *Événements. Réalisations d'un événement*

Lorsque l'ensemble Ω des possibilités est bien déterminé, on appelle *événement* toute *partie* de cet ensemble.

> Un *événement* est *une partie* de l'ensemble *des possibilités.*

Dans l'exemple 1) ci-dessus, la phrase « la proportion d'étudiants de l'échantillon ayant fait au moins 6 ans d'an-

glais est comprise entre 50 % et 75 % (bornes incluses) » définit un événement, la partie $\{ 50, 51, 52, \ldots, 74, 75 \}$ de Ω.

De même, la phrase « cette proportion n'est pas inférieure à 63 % » définit un autre événement, la partie $\{ 63, 64, \ldots, 99, 100 \}$ de Ω.

Dans l'exemple 3) ci-dessus, on peut également considérer des événements tels que « à partir de la 7^e boule tirée, il n'y a plus que des noires » ; cet événement est la partie de Ω constituée par les 2^7 suites se terminant par cinq lettres N (pourquoi ?) ; ou l'événement « dans la suite observée, il y aura au plus une boule blanche » ; ce dernier événement est la partie de Ω constituée par les 13 suites NN...N (0 boule blanche) et BNN...N, NBN...N, ..., NN...NB (1 boule blanche — 12 suites).

— Un événement est une partie X de l'ensemble Ω des résultats possibles ; on dit que *X se réalise* (ou encore « *X est réalisée* ») lorsque le résultat effectivement obtenu est l'un des éléments de X, appartient à X.

> Une *réalisation* d'un événement : une *possibilité appartenant* à cet événement.

d) *Vocabulaire des événements*

Le mot *événement* est, dans le langage propre aux probabilistes, synonyme de *partie* de l'ensemble des possibilités ; d'autres mots du vocabulaire des parties d'un ensemble (cf. *MSH*, chap. V à VIII) ont eux aussi leurs équivalents propres au langage probabiliste. C'est ainsi que l'on dit :

— *Événement élémentaire* pour une partie ne contenant qu'un *élément* de l'ensemble Ω des possibilités. Il y a ainsi autant d'événements élémentaires que de parties de *car-*

dinal 1 dans Ω, *i.e.* que d'éléments de Ω ; dans l'exemple 2) ci-dessus, il y a 13 événements élémentaires, qui sont :

$$\{0\}, \{1\}, \ldots, \{12\}$$

— *Événement impossible* pour la *partie vide* ∅ de Ω ; en effet, la partie vide est celle qui ne contient aucun des éléments de Ω, c'est-à-dire aucun des résultats *possibles*.

— *Événement certain* pour la partie pleine (Ω tout entier) de Ω ; en effet, puisque Ω est l'ensemble de *toutes* les possibilités, c'est sûrement l'une d'entre elles (*i.e.* un élément de Ω) qui se réalisera ; on est certain, par la définition même de Ω, que le résultat sera l'un de ses éléments.

— *Événements incompatibles* (ou *mutuellement exclusifs*) pour deux *parties disjointes* (c'est-à-dire sans élément commun) de Ω ; dans l'exemple 2) ci-dessus, les deux événements :

X : « le nombre de blanches tirées est au plus 5 »

et : Y : « le nombre de noires tirées est au plus 3 »

sont incompatibles. Ce sont en effet les deux parties

$$X = \{0, 1, 2, \ldots, 5\}$$

et $$Y = \{10, 11, 12\} \text{ de } \Omega = \{0, 1, \ldots, 12\}.$$

Deux événements incompatibles sont deux événements qui ne peuvent être réalisés ensemble ; aucun des résultats possibles n'est à la fois dans les deux ; il est donc *impossible* (partie vide) qu'un résultat possible soit dans leur intersection : celle-ci est vide.

Deux événements X et Y sont dits *incompatibles* si comme parties de l'ensemble Ω des possibilités, ils sont *disjoints* : $X \cap Y = \varnothing$.

Tableau récapitulatif de la correspondance entre les deux vocabulaires

Parties d'ensemble	Probabilités
Elément de Ω	Une possibilité, un résultat possible.
Partie de Ω	Un événement.
Partie de cardinal 1	Evénement élémentaire.
Partie vide	Evénement impossible.
Partie pleine	Evénement certain.
Parties disjointes	Evénements incompatibles.

e) *Simplexe et algèbre des événements*

Chaque événement, pour un ensemble Ω *donné* des possibilités, étant une partie de Ω, l'ensemble des événements est l'ensemble $P(\Omega)$ des parties de Ω ; lorsqu'il y a n possibilités (Ω a n éléments), il y a ainsi 2^n événements.

Ces événements s'organisent les uns par rapport aux autres par la *relation d'inclusion* ($X \subset Y$: tout élément de X est élément de Y) selon le *simplexe* des parties de Ω (cf. *MSH*, chap. VI). Lorsque X est inclus dans Y, les probabilistes disent parfois que l'événement X *implique* l'événement Y ; il y a en effet implication en ce sens que chaque fois que l'événement X est réalisé, c'est-à-dire que le résultat est un élément de X, alors Y l'est aussi : ce résultat étant élément de X est aussi élément de Y, qui contient X.

A ce point, l'étudiant est invité à revoir (*MSH*, chap. VI) l'organisation de $P(\Omega)$ par la relation d'inclusion, le simplexe ; on utilisera en effet dans la suite le simplexe des événements.

D'autre part, les opérations étudiées (*MSH*, chap. VIII)

entre parties d'un ensemble sont d'usage courant et constituent dans le présent contexte le *calcul des possibilités* (ou *algèbre des événements*).

A chaque événement X, on peut faire correspondre l'*événement complémentaire* X', constitué par la partie de Ω complémentaire de la partie X.

A deux événements X et Y, on fait correspondre :

— leur *intersection* $X \cap Y$, qui s'interprète comme l'événement dans lequel X et Y sont tous les deux réalisés ;

— leur *union* $X \cup Y$, qui s'interprète comme l'événement dans lequel soit X, soit Y (éventuellement les deux) est réalisé.

L'*algèbre des événements*, c'est l'algèbre des parties de Ω pour les opérations de complémentation, d'union et d'intersection, avec toutes les règles de calcul de cette algèbre (cf. *MSH*, chap. VIII).

N.B. — La phase, préalable à tout calcul des probabilités, et *a fortiori* à toute inférence statistique, dont on a parlé dans ce paragraphe, est *absolument essentielle.* Ses trois étapes principales sont :

1. *Énumérer* l'ensemble de toutes les possibilités : *l'ensemble* Ω.
2. *Organiser* les parties de Ω, ou *événements*, de façon à *situer* les uns par rapport aux autres les événements possibles : le *simplexe des parties de* Ω.
3. *Calculer* (« qualitativement ») sur les événements possibles, par intersection, union, complémentation : *l'algèbre des événements.*

2. Mesures de probabilité

a) *Mesures de probabilité sur un ensemble*

Calcul des probabilités : le bon sens mis en calcul, disait Joseph Bertrand ; le bon sens doit s'entendre ici comme l'organisation, et l'algèbre, très naturelles, sur les événements possibles (paragraphe précédent) ; et le calcul comme le calcul numérique, le calcul arithmétique.

On va donc affecter à chaque événement *possible* un nombre, sa probabilité (1) ; on ne va pas le faire n'importe comment, mais de façon à respecter, autant que faire se peut, le « bon sens » c'est-à-dire l'organisation des événements et leur algèbre.

Faire correspondre à *chaque* événement X, c'est-à-dire à chaque partie X de l'ensemble Ω des possibilités, *un* nombre et un *seul*, sa probabilité, notée $p(X)$, c'est définir *une application* de l'ensemble $P(\Omega)$ des parties de Ω dans l'ensemble $\mathbf{R}$ des nombres réels (souvent, on prendra pour probabilités des nombres rationnels ; en ce cas, l'ensemble d'arrivée est l'ensemble $\mathbf{Q}$ des rationnels). Désignons par p cette application :

$$p : P(\Omega) \to \mathbf{R}$$

La probabilité $p(X)$ d'un événement X, c'est *l'image*, dans l'application p, de X.

$$X \xrightarrow{p} p(X)$$

Toute application de $P(\Omega)$ dans $\mathbf{R}$ ne définit pas une probabilisation de Ω ; seules sont des *mesures de probabilité sur* Ω celles qui satisfont aux conditions de « bon sens » indiquées ci-dessous :

1) La probabilité d'un événement est sensée mesurer son « degré de vraisemblance » ; ce degré, on le mesure par un nombre positif ou nul.

(1) C'est une façon d'exprimer (numériquement) *l'opinion* que l'on a sur les diverses possibilités.

$$(1) \quad \forall\, X \subset \Omega, p(X) \geqslant 0$$

La probabilité d'un événement est toujours positive ou nulle : c'est la condition de *positivité.*

2) Par convention, cette mesure des degrés de vraisemblance se fait sur une échelle allant de 0 à 1 ; la probabilité 1, degré le plus élevé de l'échelle, sera prise pour l'événement *certain*, Ω (partie pleine).

$$(2) \quad p(\Omega) = 1$$

L'événement certain a pour probabilité 1 : c'est une convention d'*échelle* (1).

3) Les deux conditions précédentes ne faisaient que préciser l'échelle des nombres mesurant les probabilités : échelle de 0 à 1 ; ces conditions sont largement conventionnelles et ne mettent pas en jeu l'algèbre.

Il nous faut maintenant une condition qui mette en rapport la structure de *départ* : $\mathrm{P}(\Omega)$, avec les opérations, $\cap$, $\cup$, complémentation, et la structure *d'arrivée* : $\mathbf{R}$, avec les opérations $+$, $\times$, etc., de l'arithmétique (cf. chap. I).

Lorsque l'ensemble Ω des possibilités est fixé une fois pour toutes, il suffit d'une seule condition : lorsque deux événements X et Y sont *incompatibles* (lorsque les parties X et Y sont disjointes), alors la probabilité de leur *union* est la *somme* de leurs probabilités :

$$(3) \quad X \cap Y = \varnothing \Rightarrow p(X \cup Y) = p(X) + p(Y)$$

L'union des deux événements *incompatibles* a pour probabilité la *somme* des probabilités de ces événements : c'est la condition d'*additivité.*

(1) Cette convention s'exprime souvent en disant que la mesure p est *normée.*

N.B. — 1) La condition (3), d'*additivité* des mesures de probabilité, est tout à fait fondamentale, puisque c'est elle qui indique comment l'algèbre des événements possibles, par $\cap$ et $\cup$, se transforme, par la mesure p de probabilité, en *calcul numérique par addition*, c'est-à-dire dans la structure de groupe abélien ordonné $(\mathbf{R}, +, \geqslant)$ (cf. chap. I).

2) Additionner les probabilités de deux événements *incompatibles* pour obtenir la probabilité de leur union est très naturel ; par contre, *une telle addition n'a plus aucun sens* en général *si les événements X et Y ne sont pas disjoints ;* on s'en convaincra aisément en voyant ce qui se passerait à la limite, lorsque X et Y sont confondus ; on a alors :

$$p(X \cup Y) = p(X \cup X) = p(X) \qquad \text{(puisque } X \cup X = X\text{)}$$
$$p(X) + p(Y) = p(X) + p(X) = 2\,p(X)$$

Poser en ce cas $p(X \cup Y) = p(X) + p(Y)$ reviendrait ainsi à écrire $1 = 2$ dans l'ensemble des nombres entiers naturels.

Définition

On appelle *mesure de probabilité* sur un ensemble Ω (fini) toute application $p : \mathrm{P}(\Omega) \to \mathbf{R}$ de l'ensemble des parties de Ω dans l'ensemble des nombres réels satisfaisant aux trois conditions :

(1) $\forall X \subset \Omega \qquad p(X) \geqslant 0 \qquad$ (positivité)

(2) $p(\Omega) = 1 \qquad$ (convention d'échelle)

(3) $X \cap Y = \varnothing \Rightarrow p(X \cup Y) = p(X) + p(Y) \qquad$ (additivité)

Ces trois conditions sont fréquemment appelées dans la littérature *axiomes de Kolmogoroff* (des mesures de probabilités).

b) *Propriétés élémentaires des mesures de probabilité*

Les conditions (1), (2) et (3) ont un certain nombre de conséquences immédiates qui permettent de préciser le rapport entre « bon sens » et « calcul » numérique : des règles de *calcul* des probabilités.

1. *Un événement impossible a une probabilité nulle.*

$$p(\varnothing) = 0$$

En effet, considérons un événement X et l'événement impossible $\varnothing$; on a : $X \cap \varnothing = \varnothing$. Donc on peut appliquer la condition 3) d'additivité :

$$p(X \cup \varnothing) = p(X) + p(\varnothing)$$

Or :

$$X \cup \varnothing = X$$

donc :

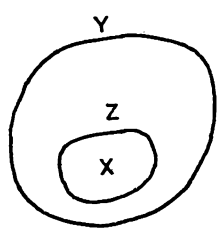
$$p(X) = p(X) + p(\varnothing) ; \qquad \text{d'où : } p(\varnothing) = 0$$

2. *Si l'événement X est inclus dans l'événement Y, la probabilité de X est inférieure ou égale à celle de Y.*

$$X \subset Y \Rightarrow p(X) \leqslant p(Y)$$

Considérons $X \subset Y$, et désignons par Z la partie de Y qui n'est pas dans X (1) ; on a par définition :

$X \cap Z = \varnothing$ et $X \cup Z = Y$



(1) $Z = Y - X = Y \cap X'$ (cf. p. 175, § c).

Appliquons la condition (3) à X et Z qui sont disjoints :

$$p(X \cup Z) = p(X) + p(Z)$$

i.e. :
$$p(Y) = p(X) + p(Z)$$

Comme, d'après la condition (1), $p(Z) \geqslant 0$, on a bien ainsi :

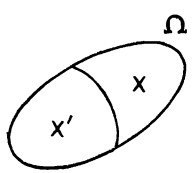
$$p(Y) \geqslant p(X)$$

Il est tout à fait naturel qu'une mesure de probabilité ait cette propriété : si l'événement Y est réalisé chaque fois que X l'est, et éventuellement d'autres fois, Y est au moins aussi probable que X.

3. *Lorsque deux événements sont complémentaires, la somme de leurs probabilités vaut 1.*

Désignons par X′ le complémentaire de X dans Ω ; on a :

$$\boxed{p(X') + p(X) = 1}$$



ou encore :

$$\boxed{p(X') = 1 - p(X)}$$

En effet, $X' \cap X = \varnothing$ et $X' \cup X = \Omega$; donc d'après (3) :

$$p(X' \cup X) = p(X) + p(X')$$

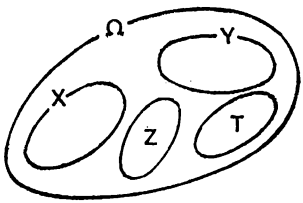
Et d'après (2) :

$$p(X' \cup X) = p(\Omega) = 1$$

4. *Si plusieurs événements sont deux à deux incompatibles la probabilité de leur union est la somme des probabilités de chacun d'eux.*

Soit X, Y, Z, T par exemple quatre tels événements. X et Y sont disjoints ($X \cap Y = \varnothing$), donc :

$$p(X \cup Y) = p(X) + p(Y)$$



Comme $X \cap Z = Y \cap Z = \varnothing$, alors $(X \cup Y) \cap Z = \varnothing$.

Démonstration :

$$(X \cup Y) \cap Z = (X \cap Z) \cup (Y \cap Z) = \varnothing \cup \varnothing = \varnothing$$

Donc, en appliquant (3) à $(X \cup Y)$ et Z qui sont disjoints :

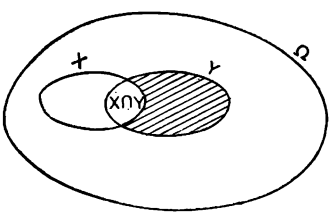
$$p(X \cup Y \cup Z) = p(X \cup Y) + p(Z) = p(X) + p(Y) + p(Z)$$

On montre ensuite de même que $(X \cup Y \cup Z)$ et T sont deux parties disjointes, et en leur appliquant (3) :

$$p(X \cup Y \cup Z \cup T) = p(X) + p(Y) + p(Z) + p(T)$$

c) *Identité des mesures*

Lorsque deux parties X et Y de Ω ne sont plus disjointes, il y a une relation appelée ici *identité des mesures*, entre la probabilité de X, celle de Y, celle de $X \cup Y$ et celle de $X \cap Y$.



Soit Z la partie de Y qui n'est pas dans X, ou *différence* entre Y et X, notée souvent $(Y - X)$, et qui par définition est

l'intersection de Y et du complémentaire X' de X dans Ω (partie hachurée de la figure).

$$Z = Y - X = Y \cap X'$$

On a $X \cap Z = \varnothing$ par définition, et $X \cup Z = X \cup Y$.
Donc, d'après (3) appliquée à X et Z (disjointes) :

(a) $$p(X \cup Y) = p(X \cup Z) = p(X) + p(Z)$$

Mais $X \cap Y$ et Z sont également disjointes, et leur union est par définition Y ; donc, (3) appliquée à $X \cap Y$ et à Z, fournit :

(b) $$p(Y) = p((X \cap Y) \cup Z) = p(X \cap Y) + p(Z)$$

De *(a)* et *(b)* respectivement on tire :

$$p(Z) = p(X \cup Y) - p(X)$$
$$p(Z) = p(Y) - p(X \cap Y)$$

D'où :

$$p(X \cup Y) - p(X) = p(Y) - p(X \cap Y)$$

Ce qui s'écrit encore sous la forme :

(M) $$\boxed{p(X) + p(Y) = p(X \cup Y) + p(X \cap Y)}$$

L'identité (M), s'énonce :

> La somme des probabilités de deux parties X et Y de l'ensemble Ω est égale à la somme des probabilités de leur union et de leur intersection.

Cette identité est appelée *identité des mesures* car on appelle *mesure sur un ensemble* Ω fini toute application μ de l'ensemble $P(\Omega)$ des parties de Ω dans l'ensemble des nombres réels satisfaisant à cette identité, et à la condition $\mu(\varnothing) = 0$.

Mesure sur un ensemble Ω (fini) : application $\mu : P(\Omega) \to \mathbf{R}$ satisfaisant à :

(M) $\mu(X \cup Y) + \mu(X \cap Y) = \mu(X) + \mu(Y)$

(4) $\mu(\varnothing) = 0$

Les mesures de probabilité sont des mesures particulières, qui satisfont à deux conditions supplémentaires : être positives (condition 1, § 2-*a*) ; $p(\Omega) = 1$ (condition 2, § 2-*a*).

d) *Applications*

L'identité (M) sert, lorsque l'on connaît la probabilité de trois des quatre événements X, Y, $X \cap Y$ et $X \cup Y$, à déterminer celle du quatrième.

Exemple :

Si : $p(X) = 1/2, p(X \cup Y) = 3/4, p(X \cap Y) = 3/8$, alors :

$$p(Y) = p(X \cup Y) + p(X \cap Y) - p(X)$$
$$= 3/4 + 3/8 - 1/2 = 5/8$$

C'est ainsi que de l'identité (M) on déduit, par exemple, comment calculer la probabilité d'une intersection $X \cap Y$, lorsque l'on connaît déjà celles de X, de Y et l'union $X \cup Y$:

$$p(X \cap Y) = p(X) + p(Y) - p(X \cup Y)$$

Exemple :

Si l'on sait que $p(X) = 1/3, p(Y) = 1/4$ et $p(X \cup Y) = 1/2$, alors :

$$p(X \cap Y) = 1/3 + 1/4 - 1/2 = 4/12 + 3/12 - 6/12 = 1/12$$

Par contre, cela n'a pas de sens de vouloir calculer la probabilité de $X \cap Y$ si l'on ne connaît que celle de X et celle de Y. Si l'on sait simplement que $p(X) = 1/3$ et $p(Y) = 1/4$, $p(X \cap Y)$ peut être n'importe quel nombre compris entre 0 et la plus petite (ici 1/4) de ces deux probabilités.

En effet, on a, d'après (M), et puisque $p(X \cap Y) \geqslant 0$:

$$\boxed{p(X \cup Y) \leqslant p(X) + p(Y)}$$

Cette inégalité est une égalité lorsque $X \cap Y = \varnothing$. D'autre part, puisque $X \cup Y \supset X$ et $X \cup Y \supset Y$,

$$\boxed{p(X \cup Y) \geqslant \max(p(X), p(Y))}$$

La probabilité de l'union de deux parties est toujours comprise entre la plus grande des deux probabilités de ces parties et la somme de leurs probabilités.

*3. Identité des mesures différences et axiomes des probabilités

On a vu que la condition (l'axiome) (3) des mesures de probabilité avait entre autres conséquences :

(4) $$p(\varnothing) = 0$$

et

(M) $$p(A \cup B) + p(A \cap B) = p(A) + p(B)$$

Inversement, les deux relations (4) et (M) impliquent (3) : si en effet $A \cap B = \varnothing$, en vertu de (4), (M) se réduit à

$$(3) \qquad p(A \cup B) = p(A) + p(B) \qquad (A \cap B = \varnothing)$$

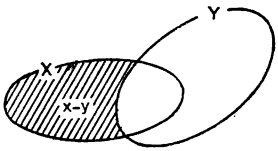
On peut donc remplacer le système de conditions (1), (2) et (3) définissant les mesures de probabilité par le système :

(1)	$\forall X \subset \Omega, \quad p(X) \geqslant 0$
(2)	$p(\Omega) = 1$
(4)	$p(\varnothing) = 0$
(M)	$p(A \cup B) + p(A \cap B) = p(A) + p(B)$

L'intérêt de cette nouvelle présentation provient de ce que les conditions ainsi écrites sont symétriques par rapport aux opérations $\cap$ et $\cup$ de l'algèbre des parties de Ω, ne privilégient aucune d'entre elles ; ceci est dans la logique de l'algèbre des parties de Ω, et du principe de *dualité* dans cette algèbre (cf. *MSH*, chap. VIII, § 3).

Une mesure de probabilité p applique la structure $(\mathbf{P}(\Omega), \cap, \cup, \complement)$ des parties de Ω, structure *d'algèbre de Boole* (*MSH*, chap. VIII), dans la structure $(\mathbf{R}, +, \geqslant)$ des nombres réels, ordonnés, et munis de l'addition, structure de *groupe abélien totalement ordonné* (chap. I, § 4). On sait que ces deux structures ont des ressemblances, mais aussi des dissemblances. Certaines propriétés de l'addition des nombres (commutativité, associativité) sont vérifiées par l'union et l'intersection, mais par contre d'autres propriétés ne sont vérifiées que par l'une d'entre elles : l'idempotence ($X \cup X = X$) de l'union et de l'intersection n'est pas vraie pour l'addition des nombres.

L'application $p : \mathbf{P}(\Omega) \to \mathbf{R}$ ne peut donc entièrement « transporter » la structure de départ dans celle d'arrivée ; il y a néanmoins une opération entre parties d'ensemble



qui est transportée par l'application p. C'est la *différence* entre parties d'ensemble, lorsque l'une de ces parties contient l'autre.

En effet, $X - Y$, qui est la partie de X qui n'est pas dans Y (hachurée dans les figures ci-contre), est disjointe de Y. Si $X \supset Y$, $(X - Y) \cup Y = X$; donc, d'après l'additivité :

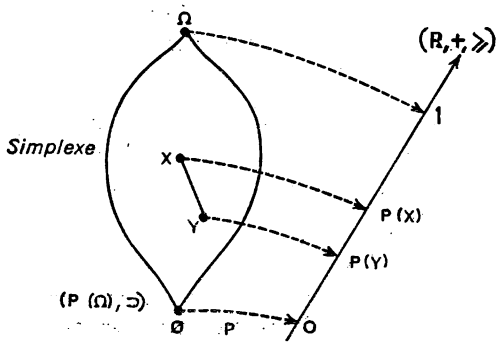
$$p(X - Y) + p(Y) = p(X)$$

Soit :

(D) $\boxed{X \supset Y \Rightarrow p(X - Y) = p(X) - p(Y)}$

Cette condition (D) implique à son tour l'additivité (3) (démontrez-le). Elle lui est *équivalente.*

Pour bien voir ce que signifie le système d'axiomes (1), (2), (M), (4), considérons $P(\Omega)$ partiellement ordonné par la relation d'inclusion (le simplexe), et le groupe abélien totalement ordonné $(\mathbf{R}, +, \geqslant)$.



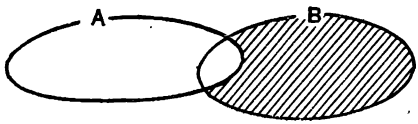
Ce qu'expriment (1), (2) et (4), c'est que p envoie le « pôle inférieur » $\varnothing$ du simplexe $(P(\Omega), \supset)$ sur 0, le « pôle supérieur » Ω sur 1, et toute autre partie X, c'est-à-dire tout autre sommet du simplexe, entre 0 et 1. Si $X \supset Y$, alors $p(X) \geqslant p(Y)$.

L'ordre d'inclusion $\supset$ est transformé, par p, en l'ordre des nombres $\geqslant$.

Mais pour tout couple A, B de parties, on a d'autre part :

$$A \cup B - A = B - A \cap B$$

(partie hachurée de la figure).



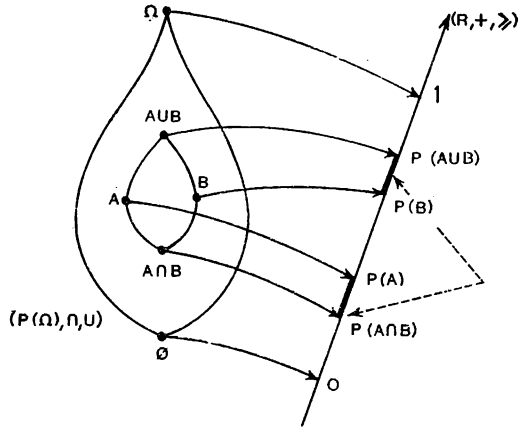
Si donc on suppose que p satisfait à (3), c'est-à-dire à la condition (D) qui lui est équivalente, p satisfait alors à la condition (M), puisque les deux différences *ensemblistes* égales $(A \cup B - A)$ et $(B - A \cap B)$ doivent avoir pour images les deux différences *numériques* égales $(p(A \cup B) - p(A))$ et $(p(B) - p(A \cap B))$ (en effet $A \cup B \supset A$, et $B \supset A \cap B$) :

$$p(A \cup B) - p(A) = p(B) - p(A \cap B)$$

Ce qui se réécrit :

$$\text{(M)} \qquad p(A \cup B) + p(A \cap B) = p(B) + p(A)$$

On voit ainsi d'une autre façon comment (M) est une conséquence de la condition (D), c'est-à-dire de (3).



Sur le diagramme ci-dessus du simplexe, on a exprimé symboliquement ce que signifie la condition (M).

> Deux intervalles *égaux*, *images* de différences d'ensemble *égales*.

4. Distributions de probabilités et construction des mesures de probabilité

a) *Probabilités des événements élémentaires*

— L'une des conséquences de la condition (3) d'additivité est que si plusieurs parties de Ω sont disjointes, la probabilité de leur union est la somme de leurs probabilités ; en particulier, deux parties réduites à un seul élément (événements élémentaires) sont toujours disjointes. Une partie X de Ω aura donc pour *probabilité la somme des probabilités des événements élémentaires qui constituent X.*

Pour fixer les idées, supposons :

$$\Omega = \{ 1, 2, 3, 4, 5 \}$$

Soit p_1 la probabilité de l'événement élémentaire $\{ 1 \}$, p_2 celle de $\{ 2 \}$, etc.

$$p_1 = p(\{ 1 \}), \quad p_2 = p(\{ 2 \}), \quad p_3 = p(\{ 3 \}),$$
$$p_4 = p(\{ 4 \}), \quad p_5 = p(\{ 5 \})$$

Quelle est la probabilité de $X = \{ 2, 4, 5 \}$?

On a :

$$X = \{ 2 \} \cup \{ 4 \} \cup \{ 5 \}$$

avec :

$$\{ 2 \} \cap \{ 4 \} = \varnothing, \quad \{ 2 \} \cap \{ 5 \} = \varnothing, \quad \{ 4 \} \cap \{ 5 \} = \varnothing$$

Donc :

$$p(X) = p(\{ 2 \}) + p(\{ 4 \}) + p(\{ 5 \})$$

$$\boxed{p(X) = p_2 + p_4 + p_5}$$

— Le raisonnement est évidemment général. Soit Ω quelconque, et ayant n éléments :

$$\Omega = \{ 1, 2, \ldots, n \}$$

Pour tout $i \in \Omega$, on désigne par p_i la probabilité de l'événement élémentaire $\{ i \}$:

$$p_i = p(\{ i \})$$

Alors, pour une partie X quelconque de Ω, $p(X)$ est la somme des p_i, lorsque i parcourt X. Ce qu'on écrit en abrégé :

$$\boxed{p(X) = \sum_{i \in X} p_i}$$

Le symbole Σ signifie ici *somme* ; Σp_i se lit : *somme des* p_i. L'indication $i \in X$ située sous le signe Σ de la somme $\left(\sum_{i \in X}\right)$ dit que cette somme est à effectuer sur tous les éléments de X.

Exemple :

$$\Omega = \{1, 2, 3, 4, 5\} \qquad X = \{2, 3, 5\}$$
$$p(X) = \sum_{i \in X} p_i = \sum_{i \in \{2,3,5\}} p_i = p_2 + p_3 + p_5$$

La probabilité d'un événement (d'une partie de Ω) est la somme des probabilités des événements élémentaires lui appartenant :

$$p(X) = \sum_{i \in X} p_i$$

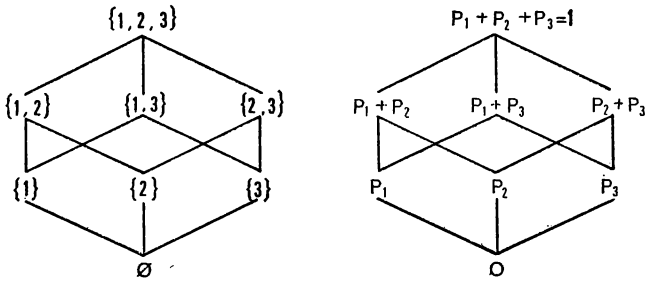
b) *Détermination d'une mesure de probabilité par les probabilités des événements élémentaires*

— Par conséquent, il suffit, pour pouvoir *calculer* la probabilité de tout événement, de connaître la probabilité de chacun des événements élémentaires. Si par exemple $\Omega = \{1, 2, 3\}$, il y a trois événements élémentaires ; désignons par p_1, p_2, p_3 respectivement leurs probabilités, on peut ensuite *calculer* :

$$p(\{1,2\}) = p_1 + p_2 \qquad p(\{2,3\}) = p_2 + p_3$$
$$p(\{1,3\}) = p_1 + p_3 \qquad p(\{1,2,3\}) = p_1 + p_2 + p_3$$

Comme $\{1, 2, 3\} = \Omega$, on a ici : $p_1 + p_2 + p_3 = 1$.

Comme d'autre part on sait que $p(\varnothing) = 0$, on connaît la probabilité de chacune des $2^3 = 8$ parties de Ω dès qu'on connaît celles des 3 événements élémentaires.



Si Ω a n éléments, les n probabilités $p_1, p_2, \ldots, p_n$ des n événements élémentaires $\{1\}, \{2\}, \ldots, \{n\}$, déterminent de même les probabilités de toutes les autres parties non vides de Ω ; pour chaque partie non vide, sa probabilité s'obtient en additionnant les probabilités des événements élémentaires lui appartenant.

La connaissance des probabilités des événements élémentaires détermine entièrement une mesure de probabilité.

c) *Distributions de probabilité et mesures*

Les probabilités des événements élémentaires ne peuvent être choisies de façon totalement arbitraire ; elles sont astreintes à des conditions, conséquences des axiomes (1), (2) et (3) des mesures de probabilité.

1° Chaque événement élémentaire étant un événement, sa probabilité doit être positive ou nulle d'après (1) : $\forall X \subset \Omega$, $p(X) \geqslant 0$. On a donc :

$$\forall i \in \Omega : p_i = p(\{i\}) \geqslant 0$$

Chacune des probabilités des événements élémentaires est positive ou nulle.

$$\forall i \in \Omega \qquad p_i \geqslant 0$$

2° Ω tout entier, événement certain, a pour probabilité 1 d'après (2) ; mais sa probabilité est la somme de toutes les probabilités élémentaires. Par conséquent :

$$\sum_{i \in \Omega} p_i = p_1 + p_2 + \ldots + p_n = 1$$

La somme des probabilités des événements élémentaires vaut 1 :

$$p_1 + p_2 + \ldots + p_n = 1$$

La suite $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ des probabilités élémentaires constitue *une distribution de probabilité* sur Ω.

Distribution de probabilité sur Ω : La suite des probabilités des événements élémentaires de Ω ; si Ω a n élément, n nombres $p_1, p_2, \ldots, p_n$ tous positifs ou nuls, et de somme unité :

$$\mathbf{p} = (p_1, p_2, \ldots, p_n)$$
$$\forall i \qquad p_i \geqslant 0$$
$$p_1 + p_2 + \ldots + p_n = 1$$

Il y a ainsi autant de distributions de probabilité sur Ω que de systèmes de n nombres positifs et de somme unité ; chaque *distribution de probabilité* détermine *une mesure de*

probabilité dans laquelle la probabilité d'une partie (non vide) de Ω est la somme des probabilités des événements élémentaires de cette partie.

Chaque distribution de probabilité :

$$\mathbf{p} = (p_1, p_2, \ldots, p_n)$$

détermine une mesure de probabilité p dans laquelle :

$$\forall X \subset \Omega \qquad p(X) = \sum_{i \in X} p_i$$

Exemples :

1) $\Omega = \{1, 2, 3, 4\}$.

$\mathbf{p} = (1/4, 1/8, 1/2, 1/8)$ est une distribution de probabilité pour laquelle :

$$p_1 = 1/4 \qquad p_2 = 1/8 \qquad p_3 = 1/2 \qquad p_4 = 1/8$$

On a bien :

$$p_1 + p_2 + p_3 + p_4 = 1/4 + 1/8 + 1/2 + 1/8 = 1$$

Dans la mesure de probabilité qu'elle détermine, on a, pour les autres parties de Ω :

$$p(\varnothing) = 0 \qquad p(\{1, 2\}) = p_1 + p_2 = 1/4 + 1/8 = 3/8$$
$$p(\{2, 3, 4\}) = p_2 + p_3 + p_4 = 1/8 + 1/2 + 1/8 = 3/4$$
$$p(\Omega) = p_1 + p_2 + p_3 + p_4 = 1, \text{ etc.}$$

Exercice :

Déterminez la probabilité de chacune de 16 parties de Ω. Reportez-les sur le diagramme en treillis du simplexe S_4, comme on l'a fait pour S_3, § 4-*b*.

2) Pour le même ensemble $\Omega = \{ 1, 2, 3, 4 \}$:

$$\mathbf{p} = (1/3, 0, 1/2, 1/6)$$

est une autre distribution de probabilité :

$$p_1 = 1/3 \qquad p_2 = 0 \qquad p_3 = 1/2 \qquad p_4 = 1/6$$
$$p_1 + p_2 + p_3 + p_4 = 1$$

Cette fois-ci, on a :

$$p(\{ 1, 2 \}) = 1/3 + 0 = 1/3$$
$$p(\{ 2, 3, 4 \}) = 0 + 1/2 + 1/6 = 2/3 \qquad \text{etc.}$$

Exercice :

Déterminez la probabilité des 16 parties de Ω. Reportez-les sur le diagramme en treillis de S_4.

3) Toujours pour $\Omega = \{ 1, 2, 3, 4 \}$:

$$\mathbf{p} = (1/4, 1/4, 1/4, 1/4)$$

est une distribution de probabilité dite *uniforme* ; la mesure qu'elle détermine est la *mesure uniforme* sur Ω. Dans cette mesure, chaque partie a pour probabilité le quotient de son effectif (son nombre d'éléments) par le nombre d'éléments de Ω.

$$p(\{ 1, 2 \}) = 1/4 + 1/4 = 2/4 = 1/2 \qquad |\{ 1, 2 \}| = 2$$

de même :

$$p(\{ 3, 4 \}) = 1/2$$
$$p(\{ 1, 2, 3 \}) = p(\{ 2, 3, 4 \}) = 3/4$$

4) Supposons $\Omega = \{ 1, 2 \}$ réduit à 2 éléments ; ici, il n'y a que deux probabilités élémentaires :

$$p_1 = p(\{ 1 \}) \qquad p_2 = p(\{ 2 \})$$
$$p_1 + p_2 = 1$$

Il suffit de connaître le seul nombre p_1 pour connaître aussi $p_2 = 1 - p_1$; p_1 peut être *n'importe quel nombre compris entre* 0 et 1.

En particulier :

— si $p_1 = 0$, on a $p_2 = 1$; cela signifie que l'événement $\{1\}$ est considéré comme impossible, et l'événement $\{2\}$ comme certain ;

— si $0 < p_1 < 1/2$, alors $1 > p_2 > 1/2$, et cela signifie que $\{1\}$ est considéré comme *moins probable* que l'événement $\{2\}$;

— si $p_1 = 1/2$, alors $p_2 = 1/2$; on a la distribution uniforme, $\{1\}$ et $\{2\}$ sont dits *équiprobables* (ils ont même probabilité).

d) *Distributions et mesures uniformes (équiprobabilité)*

Parmi toutes les distributions de probabilité sur un ensemble $\Omega = \{1, 2, \ldots, n\}$, celle pour laquelle tous les événements élémentaires sont *équiprobables*, c'est-à-dire ont même probabilité, joue un rôle particulièrement important en inférence statistique.

Dans ce cas on a :

$$p_1 = p_2 = p_3 = \ldots = p_n$$

Comme la somme des p_i vaut 1, chacun est donc égal à $1/n$, où : $n = |\Omega|$.

$$\forall_i \qquad p_i = 1/n$$

$\mathbf{p} = (1/n, 1/n, \ldots, 1/n)$ est appelée *distribution uniforme* (ou *équiprobable*) sur Ω ; la mesure de probabilité qu'elle détermine est la *mesure uniforme*.

Distribution (et mesure) uniforme : équiprobabilité des événements élémentaires.

$$p_1 = p_2 = \ldots = p_n = 1/n$$

La probabilité de chaque partie X de Ω est alors le rapport de son cardinal (son nombre d'éléments) $|X|$ à celui de n de Ω :

$$\forall X \subset \Omega,\ p(X) = \frac{|X|}{n} = \frac{|X|}{|\Omega|}$$

Toutes les parties de *même cardinal* ont donc même probabilité. Toutes les parties ayant 2 éléments (cardinal 2) ont pour probabilité $2/n$; toutes celles qui ont 3 éléments ont pour probabilité $3/n$; etc.

On verra dans le chapitre suivant comment des distributions usuelles en inférence statistique (distribution binomiale, distribution hypergéométrique) sont construites à partir des distributions uniformes.

Récapitulation

Les possibilités

L'ensemble $\Omega = \{1, 2, \ldots, n\}$ des possibilités

↓

Un *événement* : une partie de Ω

↓

Organiser les événements : le simplexe $(P(\Omega), \supset)$

↓

Algèbre des événements (calcul des possibilités) : l'algèbre des parties de Ω.

$\cap$, $\cup$, $\square'$, $\varnothing$, Ω

Les probabilités

Une *mesure de probabilité* sur Ω :

$$p : \mathrm{P}(\Omega) \to \mathbf{R}$$

(1) $\forall X,\ p(X) \geqslant 0$ positivité ;

(2) $p(\Omega) = 1$ convention d'échelle ;

(3) $X \cap Y = \varnothing \Rightarrow p(X \cup Y) = p(X) + p(Y)$ additivité.

↓

Toute mesure de probabilité p satisfait à *l'identité des mesures* :

(M) $p(A \cup B) + p(A \cap B) = p(A) + p(B)$

↓

Une *distribution de probabilité* sur Ω :

$$\mathbf{p} = (p_1, p_2, \ldots, p_n)$$

$i \in \Omega$, $p_i = p(\{i\})$: probabilité de l'événement élémentaire $\{i\}$.

1) $\forall i,\ p_i \geqslant 0$

2) $p_1 + p_2 + \ldots + p_n = 1$

↓

Chaque *distribution de probabilité* détermine une *mesure de probabilité* par :

$$\forall X \subset \Omega,\ p(X) = \sum_{i \in X} p_i$$

↓

Distribution uniforme :

Tous les événements élémentaires ont même probabilité.

$$p_1 = p_2 = \ldots = p_n = 1/n$$

CHAPITRE VII

ALÉAS

1. ALÉAS

a) *Définition. Égalité*

On appelle ici *aléa* la donnée d'un ensemble Ω (fini) de possibilités et d'une mesure de probabilité p sur Ω ; ou, ceci revient au même, d'un ensemble Ω et d'une distribution de probabilité $\mathbf{p}$ sur Ω (chap. VI, § 4).

Un *aléa* est donc un *couple* (Ω, p), ou $(\Omega, \mathbf{p})$; Ω est appelé ensemble de *définition* de l'aléa, p sa mesure de probabilité, $\mathbf{p}$ sa distribution de probabilité.

Deux aléas sont *égaux* si et seulement si ils ont même ensemble de définition, et même distribution de probabilité sur cet ensemble.

> *Un aléa :* un couple $(\Omega, \mathbf{p})$, où Ω est un ensemble fini, et $\mathbf{p}$ une distribution de probabilité sur Ω.

Synonyme fréquent : variable aléatoire.

Exemple :

Les aléas de Bernoulli. — Un exemple particulièrement simple d'aléa est fourni par le cas où Ω se réduit à deux éléments, notés 0 et 1 (cf. chap. VI, § 4-*b*-4).

$$\Omega = \{0, 1\}$$

Dans les interprétations, Ω pourra par exemple être l'ensemble { Pile, Face }, ou bien 0 et 1 signifieront absence ou présence de tel caractère, etc.

Si l'on se donne une distribution de probabilité $\mathbf{p} = (p_0, p_1)$ sur Ω, on définit un *aléa* $(\Omega, \mathbf{p})$, appelé aléa de Bernoulli ; comme $p_0 + p_1 = 1$, cet aléa est entièrement défini par la donnée, outre Ω, de la valeur $\theta = p_1$ $(0 \leqslant \theta \leqslant 1)$.

On a alors $p_0 = 1 - \theta$.

A chaque valeur de θ correspond un aléa de Bernoulli ; θ est le *paramètre* dont dépend l'aléa de Bernoulli. A deux valeurs θ et θ' distinctes correspondent deux aléas de Bernoulli distincts.

Aléa de Bernoulli de paramètre θ — couple $(\Omega, \mathbf{p})$
où :
$\Omega = \{ 0, 1 \}$ $\quad \mathbf{p} = (1 - \theta, \theta)$ $\quad 0 \leqslant \theta \leqslant 1$

Les aléas sont souvent présentés sous la forme de la table qui, à chaque élément de Ω, fait correspondre sa probabilité (la table de l'application $\mathbf{p} : \Omega \to \mathbf{R}$) :

Ω	1	2	3	...	n
$\mathbf{p}$	p_1	p_2	p_3	...	p_n

où $\Omega = \{ 1, 2, \ldots, n \}$ et $\mathbf{p} = (p_1, p_2, \ldots, p_n)$.

En particulier, on écrit ainsi pour les aléas de Bernoulli :

0	1
p_0	p_1

ou

0	1
$1 - \theta$	θ

$p_0 + p_1 = 1$ $\qquad$ $0 \leqslant \theta \leqslant 1$
$p_0 \geqslant 0, p_1 \geqslant 0$

b) *Aléas finis sur un ensemble infini*

On a considéré jusqu'ici que l'ensemble Ω des possibilités était *fini* ; dans la suite, on est amené à considérer des aléas dont l'ensemble Ω de *définition est infini* (pratiquement, Ω est un ensemble de nombres), mais tels que seul un nombre *fini* d'éléments de Ω aient une probabilité non nulle.

On dit encore dans ce cas que la distribution $\mathbf{p}$ est finie sur Ω, et que l'aléa $(\Omega, \mathbf{p})$ est fini.

Aléa fini : aléa $(\Omega, \mathbf{p})$ tel que $p_i > 0$ si et seulement si $i \in A$, où A est une partie *finie* de Ω.

La partie *finie* A dont les éléments ont une probabilité non nulle est souvent appelée le *support* de la mesure de probabilité p (ou de la distribution de probabilité $\mathbf{p}$).

Exemple :

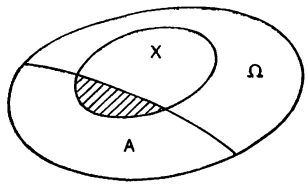
Un aléa de Bernoulli peut toujours être considéré comme un aléa ayant N pour ensemble de définition, et pour support la partie $A = \{0, 1\}$ constituée par les deux premiers entiers naturels. Ce point de vue est systématiquement exploité dans la suite (chap. IX); s'il est adopté, l'aléa de Bernoulli de paramètre θ est le couple $(\mathbf{N}, \pi)$, où π est, sur N, la distribution de probabilité :

$$\pi = (1 - \theta, \theta, 0, 0, 0, \ldots, 0, \ldots)$$

Si $(\Omega, \mathbf{p})$ est fini de support A, on définit la mesure p de probabilité sur Ω (infini) en attribuant par convention une probabilité *nulle* à toute partie *contenue dans la partie*

complémentaire de A (1). Autrement dit, on respecte toutes les règles (chap. précédent) du calcul des probabilités en posant pour toute partie $X \subset \Omega$:

$$p(X) = \sum_{i \in X \cap A} p_i$$



Exercice :

Vérifier que la mesure p ainsi définie à partir de la distribution **p** de support fini A satisfait bien aux règles du calcul des probabilités : positivité, condition d'échelle, additivité.

2. Aléa défini par une application

a)

Soit Ω un ensemble de possibilités, et p une mesure de probabilité sur Ω. Soit f une *application* de Ω dans un ensemble Ω' :

$$f : \Omega \to \Omega'$$

L'application f permet de définir à partir de l'*aléa* $(\Omega, \mathbf{p})$ un nouvel aléa dont l'ensemble de définition est Ω' : l'*aléa image* par f de $(\Omega, \mathbf{p})$.

Exemple :

On jette un dé rouge et un dé blanc sur une table, et on observe la face montrée par chacun des dés. L'ensemble Ω

(1) Les parties de probabilité nulle sont souvent appelées *négligeables* (par rapport à p) ou *de mesure nulle*.

des possibilités est constitué par les $36 (= 6 \times 6)$ couples (u, v) de nombres entiers de 1 à 6, où u désigne la face montrée par le dé rouge, et v celle qui est montrée par le dé vert.

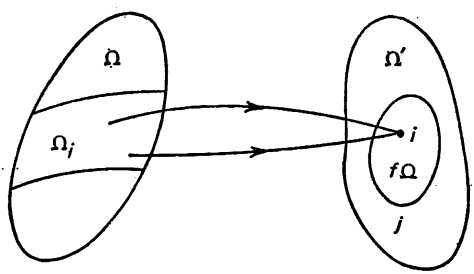
Si les dés ne sont pas pipés, on attribue à chacun des 36 événements élémentaires la même probabilité $1/36$: p est la mesure uniforme. Si à chaque résultat (couple (u, v)) on fait correspondre la somme $u + v$ des points marqués, on définit une application $f : \Omega \to \Omega'$, où Ω' est l'ensemble $\mathbf{N} = \{0, 1, 2, \ldots, n, \ldots\}$ des entiers naturels, avec :

$$f(u, v) = u + v$$

b) *Aléa image; mesure image*

La distribution de probabilité de l'aléa image, c'est la donnée pour chaque $i \in \Omega'$, ensemble d'arrivée, de la probabilité que fx soit égal à i. Plus précisément, l'application f définit dans Ω' une *partie, l'image* $f\Omega$ de Ω par f; et elle induit dans Ω une *partition* et le quotient Ω/f de Ω par la relation d'équivalence : « avoir même image par f » (cf. *MSH*, chap. I, § 3).

Si i, élément de Ω', est image d'éléments de Ω, on désigne par Ω_i la classe des éléments de Ω dont l'image est i.



$$\Omega_i = \{x : fx = i\}$$

En désignant par f^{-1} la relation de Ω' vers Ω réciproque de l'application f (f^{-1} n'est pas une application en général : certains éléments de Ω' ne sont image d'aucun élément de Ω, et certains le sont de plusieurs éléments de Ω), ceci peut encore s'écrire : $\Omega_i = f^{-1} i$.

L'événement Ω_i a pour probabilité $p(\Omega_i)$ dans Ω. On attribue, dans Ω', cette probabilité à l'événement élémentaire $\{i\}$ de Ω', et l'on pose :

$$p'_i = p(\Omega_i) = p(f^{-1} i)$$

Si maintenant on considère la partie complémentaire de $f\Omega$ dans Ω', on lui attribue une probabilité nulle ; tout élément j de cette partie complémentaire a ainsi une probabilité nulle. La probabilité pour que fx soit égal à j est évidemment nulle, puisque aucun événement de Ω ne peut réaliser l'égalité $fx = j$. On pose donc :

$$\text{si } j \notin f\Omega, \quad \text{alors } p'_j = 0$$

D'ailleurs, on a dans ce cas, $f^{-1} j = \varnothing$; donc :

$$p'_j = p(f^{-1} j) = p(\varnothing) = 0$$

La formule $p'_i = p(f^{-1} i)$ définit donc $\mathbf{p}'$, pour tout $i \in \Omega'$.

On définit ainsi dans Ω' une distribution de probabilité $\mathbf{p}'$, *la distribution de probabilité de l'aléa image de* $(\Omega, \mathbf{p})$ *par* f. La mesure de probabilité p' définie par la distribution $\mathbf{p}'$ sur Ω' est souvent appelée *mesure image* de p par f. L'aléa $(\Omega', \mathbf{p}')$ image de (Ω, p) par f a pour support l'image $f\Omega$ de Ω par f, ou parfois une partie seulement de $f\Omega$: lorsque certains éléments de Ω ont une probabilité nulle. On note souvent de façon symbolique : $p' = fp$, $\mathbf{p}' = f\mathbf{p}$.

Il faut vérifier que $\mathbf{p}'$ est bien une distribution de probabilité sur Ω'. Or :

1) Tous les p'_i sont bien positifs ou nuls puisqu'ils sont soit *nuls* (si $i \notin f\Omega$), soit égaux à $p(\Omega_i) \geqslant 0$ (si $i \in f\Omega$).

2) Leur somme vaut 1, car les seuls termes non nuls de celle-ci sont ceux qui correspondent aux éléments i de $f\Omega$, support de $(\Omega', \mathbf{p}')$; c'est donc la somme des $p(\Omega_i)$; comme les parties Ω_i sont deux à deux disjointes, la somme de leurs probabilités $p(\Omega_i)$ est la probabilité de leur union ; mais leur union, c'est Ω, dont la probabilité vaut 1.

Exemple :

Reprenons l'exemple des dés. Ω' est l'ensemble $\mathbf{N}$ des entiers naturels ; l'image $f\Omega$ de Ω par f est constituée par les 11 valeurs possibles de la somme $u + v$: $f\Omega = \{2, 3, 4, \ldots, 12\}$.

Les classes Ω_i dans Ω, et la distribution de probabilité $\mathbf{p}'$ image de la distribution *uniforme* $\mathbf{p}$ sur Ω, sont données dans le tableau ci-dessous :

	$\Omega = \{(u, v)\}$	$f\Omega = \{u + v\}$	$\mathbf{p}'\left(p'_i = \frac{\lvert\Omega_i\rvert}{36}\right)$
Ω_2	(1, 1)	2	$p'_2 = 1/36$
Ω_3	(1, 2), (2, 1)	3	$p'_3 = 2/36 = 1/18$
Ω_4	(1, 3), (2, 2), (3, 1)	4	$p'_4 = 3/36 = 1/12$
Ω_5	(1, 4), (2, 3), (3, 2), (4, 1)	5	$p'_5 = 4/36 = 1/9$
Ω_6	(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)	6	$p'_6 = 5/36$
Ω_7	(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)	7	$p'_7 = 6/36 = 1/6$

Par une symétrie évidente, on détermine les classes Ω_8 à Ω_{12}, et l'on a :

$p'_8 = p'_6$; $p'_9 = p'_5$; $p'_{10} = p'_4$; $p'_{11} = p'_3$ et $p'_{12} = p'_2$

L'*aléa image* d'un aléa (Ω, p) par une application $f : \Omega \to \Omega'$, c'est l'aléa dont :

— l'ensemble de définition est l'ensemble d'arrivée Ω' de f ;

— le support est l'image $f\Omega$ de l'application f (ou une partie de $f\Omega$) ;

— la distribution de probabilité $\mathbf{p}'$ est définie par :

$$\forall i \in \Omega' \qquad p'_i = p\,(f^{-1} i)$$

La mesure p' de distribution de probabilité $\mathbf{p}'$ ainsi définie est appelée *mesure image* de p par f.

N.B. — 1. Une application $f : \Omega \to \Omega'$, où Ω est muni d'une mesure de probabilité, est fréquemment appelée une *variable aléatoire*. Mais ce terme (variable aléatoire) désigne parfois aussi ce qu'on a appelé ici un *aléa*, *i.e.* un couple (Ω, p), où p est une mesure de probabilité sur Ω. En raison de cette ambiguïté, l'expression variable aléatoire ne sera pas utilisée ici.

2. Si l'on prend pour f l'*application identique* I de Ω dans lui-même $(\forall x \in \Omega, \mathrm{I}x = x)$, l'aléa image de (Ω, p) est (Ω, p) lui-même. Ceci explique en partie l'ambiguïté sur le sens du mot « variable aléatoire » signalée ci-dessus.

3. *Construire des applications* $\mathrm{f} : \Omega \to \Omega'$, *lorsque* Ω *est muni d'une distribution de probabilité uniforme, permet d'obtenir, comme dans l'exemple* supra, *des distributions non uniformes à partir de distributions uniformes. On va maintenant examiner quelques cas particulièrement importants pour la statistique inductive dans lesquels ce principe de construction est appliqué.*

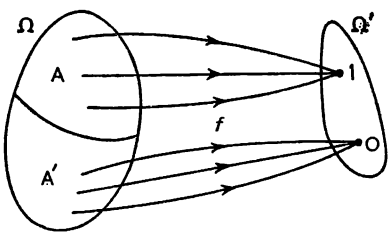
3. Aléa de Bernoulli comme aléa image

Soit E une population finie, assimilée à un ensemble ; E contient deux catégories d'individus : ceux qui présentent un certain caractère α, et qui constituent une partie A de E, et ceux qui ne le présentent pas.

Dans E, on prélève un individu : l'ensemble Ω des possibilités est identique à E si le mode de prélèvement est tel que n'importe quel élément de E puisse effectivement être choisi. Pour certains modes de prélèvement (dont la technique n'est pas étudiée ici) on peut considérer en outre que chaque élément de E a la même probabilité d'être prélevé. On a alors sur Ω (= E) une distribution **p** de probabilité uniforme.

Si maintenant on observe l'individu prélevé, ou bien il présente le caractère α, ou bien il ne le présente pas ; on définit une application $f : \Omega \to \{ 0, 1 \}$, en notant 0 l'absence de caractère α, et 1 sa présence.

L'aléa image est un aléa de Bernoulli (§ 1-*a*). La partition induite par f sur Ω a deux classes, A et sa partie complémentaire A'.



Donc : $p'_0 = p\,(A')$.
Et : $p'_1 = p\,(A)$.
Comme p est supposée uniforme sur Ω (cf. chap. VI, § 4) :

$$p'_0 = \frac{|A'|}{|\Omega|} = \frac{|A'|}{|E|} \qquad p'_1 = \frac{|A|}{|\Omega|} = \frac{|A|}{|E|}$$

p'_1, c'est la probabilité pour que l'individu prélevé possède le caractère α ; elle est égale à la proportion des individus présentant ce caractère dans la population.

L'aléa image est donc un aléa de Bernoulli dont le paramètre θ est la proportion d'individus présentant le caractère α dans la population.

Cet exemple montre que le même aléa (ici celui de Bernoulli, pour une valeur fixée du paramètre) peut être obtenu comme image d'une infinité d'autres aléas : ici, tous les aléas $(\Omega, \mathbf{p})$, où $\mathbf{p}$ est la distribution uniforme sur Ω, où Ω est fini, et où la proportion $\frac{|A|}{|\Omega|}$ a la même valeur θ.

Exemples :

— $\Omega = \{a, b, c, d, e, f\}$; $A = \{a, b\}$ fournit l'aléa de Bernoulli :

0	1
2/3	1/3

— Mais $\Omega = \{1, 2, 3, \ldots, 27\}$ $A = \{3, 6, 9, \ldots, 27\}$ fournit aussi comme aléa image :

0	1
2/3	1/3

4. Aléas hypergéométriques (tirages sans remise)

a) Si au lieu de prélever un seul élément de E on en prélève un « échantillon » de taille n, c'est-à-dire une partie de n éléments, on extrait de E une partie de n éléments ; l'ensemble Ω

des résultats possibles est celui des parties de cardinal n de E ; si E a m éléments, on a :

$$|\Omega| = \binom{m}{n}$$

$\binom{m}{n}$ désigne le nombre de parties de cardinal n dans un ensemble ayant m éléments (cf. *MSH*, chap. VII).

Bien entendu, dans la pratique de l'expérience, ceci suppose que les n individus de la population sont tous prélevés à la fois, ou bien, s'ils le sont un à un, qu'on ne réintroduise pas dans la population un individu tiré ; on a affaire à ce qu'on appelle un tirage *sans remise*, ou encore *exhaustif*. Si les conditions expérimentales sont convenables, on peut encore faire l'hypothèse d'équiprobabilité sur l'ensemble des $\binom{m}{n}$ résultats possibles ; chaque partie a une probabilité $1 \Big/ \binom{m}{n}$ d'être tirée. On a donc un aléa $(\Omega, \mathbf{p})$ où $\mathbf{p}$ est *uniforme*.

Exemple :

$$m = 10,\ n = 3\ ;\ |\Omega| = \binom{10}{3} = \frac{10!}{3!\,7!} = \frac{10.9.8}{1.2.3} = 120\ ;$$

$$\mathbf{p} = \left(\frac{1}{120}, \ldots, \frac{1}{120}\right).$$

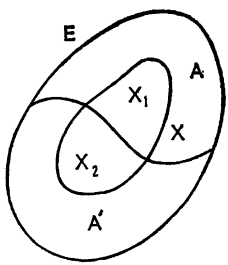
b) E est partitionné en deux classes, la partie A constituée par des individus présentant le caractère α, et la partie complémentaire.

A chacune des $\binom{m}{n}$ parties de cardinal n de E, c'est-à-dire à chaque élément de Ω, on fait correspondre le nombre des

éléments de cette partie qui présentent le caractère α ; c'est définir une application $f : \Omega \to \mathbf{N}$ (N ensemble des entiers naturels), dont l'image est :

$$f\Omega = \{0, 1, 2, \ldots, n\}$$

Pour avoir la distribution de probabilité p′ de l'aléa image, il faut déterminer, pour chaque $i \in f\Omega$, la classe $\Omega_i = f^{-1} i$; Ω_i, c'est l'ensemble des parties X de cardinal n de E dont l'intersection avec A a pour cardinal i.



$$\Omega_i = \{X : X \subset E, \quad |X| = n, \quad |X \cap A| = i\}$$

Une telle partie X est partitionnée par A et sa partie complémentaire A′ en deux parties, $X \cap A$ et $X \cap A'$, ayant respectivement i et $n - i$ éléments. Réciproquement tout couple (X_1, X_2) de parties de Ω ayant respectivement i et $(n - i)$ éléments et telles que $X_1 \subset A$ et $X_2 \subset A'$, est tel que l'union $X = X_1 \cup X_2$ est une partie de cardinal n dont i éléments présentent le caractère α (puisque i éléments appartiennent à A).

Donc Ω_i est en correspondance bijective avec l'ensemble des couples (X_1, X_2) de parties, telles que $|X_1| = i$, $|X_2| = n - i$, $X_1 \subset A$, $X_2 \subset A'$; cet ensemble de couples, c'est le produit cartésien $\mathscr{P}_1 \times \mathscr{P}_2$, où $\mathscr{P}_1$ est l'ensemble

des parties de cardinal i de A, et $\mathscr{P}_2$ l'ensemble des parties de cardinal $(n-i)$ de A'.

Supposons : $|A| = k$ $(k \leqslant m)$.

On a : $|A'| = m - k$.

Et :

$$|\Omega_i| = |\mathscr{P}_1 \times \mathscr{P}_2| = |\mathscr{P}_1| \times |\mathscr{P}_2| = \binom{k}{i}\binom{m-k}{n-i}$$

D'où finalement :

$$p'_i = p(f^{-1}i) = p(\Omega_i) = \frac{|\Omega_i|}{|\Omega|} = \frac{\binom{k}{i}\binom{m-k}{n-i}}{\binom{m}{n}}$$

Exemple :

Supposons toujours $m = 10$, $n = 3$ et en outre $|A| = k = 4$; on a : $p' = (p'_0, p'_1, p'_2, p'_3, 0, 0, \ldots)$.

$$p'_0 = \frac{\binom{4}{0}\binom{6}{3}}{\binom{10}{3}} = \frac{1 \times 20}{120} = \frac{1}{6}$$

$$p'_1 = \frac{\binom{4}{1}\binom{6}{2}}{\binom{10}{3}} = \frac{4 \times 15}{120} = \frac{1}{2}$$

$$p'_2 = \frac{\binom{4}{2}\binom{6}{1}}{\binom{10}{3}} = \frac{6 \times 6}{120} = \frac{3}{10}$$

$$p'_3 = \frac{\binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{4 \times 1}{120} = \frac{1}{30}$$

$$\mathbf{p}' = (1/6, 1/2, 3/10, 1/30, 0, 0, \ldots)$$
$$= (5/30, 15/30, 9/30, 1/30, 0, \ldots)$$

c) La distribution obtenue $\mathbf{h} = (h_0, h_1, \ldots, h_n)$ sur l'ensemble **N** des entiers naturels, a pour support la partie $\{0, 1, 2, \ldots, n\}$ de **N**, et l'on a :

$$h_i = \frac{\binom{k}{i}\binom{m-k}{n-i}}{\binom{m}{n}} \qquad (\text{pour } 0 \leqslant i \leqslant n)$$

Elle est appelée *distribution hypergéométrique* de paramètres $(m, k ; n)$; on peut la noter : $\mathbf{h}(m, k ; n)$. En effet, elle est déterminée par la donnée des trois nombres m, k et n.

On a vu que $\mathbf{h}(m, k ; n)$ peut s'interpréter en statistique comme distribution de probabilité du nombre d'éléments présentant un caractère (dichotomique) donné dans un échantillon de taille n, extrait « sans remise » d'une population d'effectif m, dont k éléments présentent le caractère α, dans l'hypothèse où tous les échantillons possibles sont équiprobables.

m : effectif de la population ;
k : nombre d'éléments présentant le caractère α ;
n : taille de l'échantillon.

$$\mathbf{h}(m, k\,;n) = (h_0, h_1, \ldots, h_i, \ldots, h_n)$$

$$h_i = \frac{\binom{k}{i}\binom{m-k}{n-i}}{\binom{m}{n}}$$

Remarque :

L'aléa (N, h) est défini sur N ; on a dit que son *support* est $\{0, 1, 2, \ldots, n\}$, si n éléments sont prélevés. En fait, si $n > k$ (si l'on prélève plus d'éléments de E qu'il n'y a d'éléments de E présentant le caractère α) le support est $\{0, 1, 2, \ldots, k\}$; d'ailleurs l'expression de h_i fournit bien dans ce cas, pour $k < i \leqslant n$, $h_i = 0$, puisque $\binom{k}{i}$ est alors nul. De même, si $m - k < n$, le support est borné par $(m - k)$.

5. Aléa de pile ou face

a) D'une population (ensemble E) ayant autant d'éléments présentant un caractère α donné que d'éléments ne le possédant pas (la proportion, dans la population, des individus présentant le caractère α est 1/2), on extrait successivement n individus ; on observe si l'individu prélevé présente ou non le caractère α, et on le remet dans la population (de sorte que le même individu peut éventuellement être observé plusieurs fois). On constitue ainsi ce qu'on appelle en statistique un échantillon par tirages « avec remise », ou « échantillon simple au hasard ».

— On lance n fois de suite une même pièce de monnaie sur une table, et l'on note à chaque jet si la pièce est tombée sur pile ou sur face.

— Un père de famille a successivement n enfants ; pour chaque enfant, on note si le sexe est féminin ou masculin. Et l'on estime qu'à chaque naissance il y a une probabilité égale (1/2) pour chacun des sexes.

Dans ces trois expériences, l'ensemble Ω des possibilités est le même ; si on désigne par $1, 2, \ldots, n$ les numéros d'ordre :

— des individus prélevés dans le premier exemple ;
— des jets d'une pièce dans le second ;
— des enfants dans le troisième ;

chaque résultat possible est une application de l'ensemble $X = \{1, 2, \ldots, n\}$ dans l'ensemble $\{0, 1\}$, où 0 note l'absence du caractère α, et 1 sa présence (premier exemple). Ω est l'ensemble des 2^n applications de X dans $\{0, 1\}$.

Exemple :

$n = 5$; il y a $2^5 = 32$ résultats possibles ; en voici quelques-uns.

00000 : aucun des éléments observés ne présente le caractère α ;
01010 : le 2^e et le 4^e des éléments observés présentent le caractère α.

Autre façon de dire : Ω est l'ensemble de 2^n parties de $X = \{1, 2, \ldots, n\}$, à chaque observation on fait correspondre la partie constituée par les numéros d'ordre des éléments (de E dans le premier exemple) présentant le caractère α.

b) On peut estimer que chacun de 2^n résultats possibles a la même probabilité ; on a alors sur Ω la distribution uniforme, dans laquelle chaque résultat a $1/2^n$ pour probabilité. On a ainsi un aléa (Ω, p).

c) Si à chaque résultat possible (élément de Ω) on fait correspondre le nombre des individus présentant le caractère α (premier exemple), le nombre de fois où pile est apparu (second exemple), le nombre de garçons parmi les enfants (troisième exemple), on définit une application :

$$f : \Omega \to \mathbf{N}$$

Le support de l'aléa image de (Ω, p) par f est l'image $f\Omega$, c'est-à-dire la partie $\{0, 1, 2, \ldots, n\}$ de $\mathbf{N}$; en effet, ou bien aucun individu prélevé ne présente le caractère α, ou un seul, ou deux, ..., ou tous.

L'aléa image obtenu a pour distribution de probabilité $\mathbf{p}' = (p'_0, p'_1, \ldots, p'_n)$, avec :

$$p' = p(\Omega_i) = p(f^{-1} i) = \frac{|\Omega_i|}{|\Omega|} = \frac{|\Omega_i|}{2^n}$$

Or Ω_i, c'est l'ensemble des parties de $\{1, 2, \ldots, n\}$ ayant i éléments ; donc $|\Omega_i| = \binom{n}{i}$ et :

$$\boxed{p'_i = 1/2^n \binom{n}{i}}$$

Exemple :

Pour $n = 5$, la distribution $\mathbf{p}' = (p'_0, p'_1, \ldots, p'_5)$ est :

	p'_0	p'_1	p'_2	p'_3	p'_4	p'_5
p'	$1/2^5 = 1/32$	5/32	10/32	10/32	5/32	1/32

Les numérateurs sont les nombres de la n^e ligne du triangle de Pascal.

6. Aléa binomial (tirages avec remise)

Si l'on reprend la question traitée au § 5, mais en supposant que la probabilité qu'un élément possède le caractère α n'est plus $1/2$ mais un nombre θ ($0 < \theta < 1$), on montre que la distribution obtenue est $\mathbf{b} = (b_0, b_1, \ldots, b_n)$, avec :

$$b_i = \binom{n}{i} \theta^i (1 - \theta)^{n-i}$$

b_i est la probabilité pour que dans un échantillon de taille n, tiré « avec remise », le nombre des éléments qui présentent le caractère α soit i.

La démonstration sera faite au chapitre VIII. L'aléa $(\mathbf{N}, \mathbf{b})$ ainsi défini est appelé *aléa binomial* de paramètres n et θ ; on peut noter $\mathbf{b}(n ; \theta)$ sa distribution, appelée *distribution binomiale*.

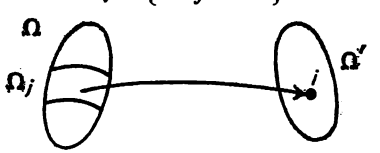
On remarque que pour $\theta = 1/2$, on a $(1 - \theta) = 1/2$, donc $\theta^i (1 - \theta)^{n-i} = 1/2^i . 1/2^{n-i} = 1/2^n$; on retrouve l'aléa de pile ou face comme cas particulier (et comme il fallait s'y attendre).

7. Aléas quotients

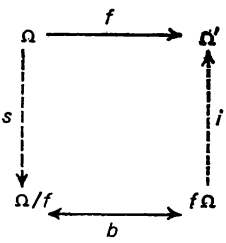
a) *Aléa quotient défini par une application*

Soit $(\Omega, \mathbf{p})$ un aléa et $f : \Omega \to \Omega'$ une application ; on a vu que l'application f induit dans son ensemble de départ une *partition* en classes Ω_i :

$$\Omega_i = \{ x : fx = i \}$$



Il y a une correspondance bijective entre l'ensemble *quotient* $\Omega/f = \{ \Omega_1, \Omega_2, \ldots, \Omega_i, \ldots \}$ et l'*image* $f\Omega = \{ 1, 2, \ldots, i, \ldots \}$. Cette correspondance est d'ailleurs mise en évidence dans la décomposition (ou factorisation) canonique de f en produit de composition d'une *surjection* $s : \Omega \to \Omega/f$ suivie d'une *bijection* $b : f/\Omega \to f\Omega$, suivie d'une *injection* $i : f\Omega \to \Omega'$, rappelée dans le diagramme ci-contre (cf. *MSH*, chap. II, § 3) :



$$f = i \circ b \circ s$$

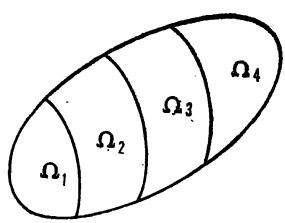
Si l'on munit l'ensemble quotient Ω/f, c'est-à-dire l'ensemble des classes de la partition induite par f, de la distribution de probabilité $\mathbf{p}'$ image :

$$\forall i \in \Omega', \quad \mathbf{p}'_i = p(\Omega_i) = p(f^{-1} i)$$

on obtient un aléa $(\Omega/f, \mathbf{p}')$ appelé ici *aléa quotient* de $(\Omega, \mathbf{p})$ par f. L'aléa quotient $(\Omega/f, \mathbf{p}')$ est d'ailleurs une « réplique » de l'aléa $(f\Omega, \mathbf{p}')$, c'est-à-dire de la restriction de l'aléa image $(\Omega', \mathbf{p}')$ à son support $f\Omega$.

b) *Aléa quotient défini par une partition*

— Comme à chaque partition π de l'ensemble Ω on peut faire correspondre une application f, celle qui applique Ω sur son quotient par cette partition, et fait correspondre à chaque $x \in \Omega$ sa classe dans la partition π, on voit que par ce procédé on peut, à partir d'un aléa $(\Omega, \mathbf{p})$, définir autant d'aléas quotient qu'il y a de partitions de Ω. Plus précisément, soit $\pi : \Omega_1, \Omega_2, \ldots, \Omega_i, \ldots$ une partition de Ω, et $\Omega/\pi = \{ \Omega_1, \Omega_2, \ldots, \Omega_i, \ldots \}$ le quotient. L'aléa quotient de $(\Omega, \mathbf{p})$ par la partition π



est l'aléa $(\Omega/\pi, \mathbf{p}')$ où $\mathbf{p}'$ est la distribution de probabilité qui à chaque classe Ω_i de π attribue la probabilité : $\mathbf{p}'_i = p(\Omega_i) = \sum_{x \in \Omega_i} \mathbf{p}_i$.

Le chapitre VIII donnera l'occasion d'en montrer des exemples.

Récapitulation

Un aléa est un couple (Ω, p) où :

Ω est un ensemble (l'ensemble de *définition* de l'aléa) ;

p est une mesure de probabilité sur Ω (la mesure de probabilité de l'aléa).

↓

Si (Ω, p) est un aléa, à chaque application :

$$f : \Omega \to \Omega'$$

correspond :

— un *aléa image* (Ω', p') ;

— un *aléa quotient* $(\Omega/f, p')$.

p' est la *mesure image* de p par f ; on peut noter : $p' = fp$.

Elle est définie par la distribution $\mathbf{p}' = (p'_1, p'_2, \ldots)$.

$$\forall i \in \Omega', \; p'_i = p(\Omega_i) = p(f^{-1} i)$$

Aléa de Bernoulli, de paramètre θ :

$$\Omega = \{0, 1\}$$
$$\mathbf{p} = (1 - \theta, \theta)$$

Aléa binomial, de paramètres n et θ :
Tirages *avec remise*.
n = taille de l'*échantillon* ;
θ = proportion *dans la population* des individus présentant le caractère étudié.
(Cas particulier, $\theta = 1/2$: pile ou face.)

$$\Omega = \{0, 1, \ldots, i, \ldots, n\}$$
$$\mathbf{p} = (p_0, p_1, \ldots, p_i, \ldots, p_n)$$

$$p_i = \binom{n}{i} \theta^i (1 - \theta)^{n-i}$$

i = nombre d'individus présentant le caractère étudié dans *l'échantillon*.

Aléa hypergéométrique de paramètres m, k, n :
Tirages *sans remise*.
m = effectif de la *population* ;
n = taille de l'*échantillon* ;
k = nombre d'individus présentant le caractère étudié dans la *population*.

$$\Omega = \{0, 1, \ldots, i, \ldots, n\}$$
$$\mathbf{p} = (p_0, p_1, \ldots, p_i, \ldots, p_n)$$

$$p_i = \frac{\binom{k}{i}\binom{m-k}{n-i}}{\binom{m}{n}}$$

i = nombre d'individus présentant le caractère étudié dans l'*échantillon*.

CHAPITRE VIII

PROBABILITÉS CONDITIONNELLES

1. MESURES DE PROBABILITÉ CONDITIONNELLES

a) *Un événement est réalisé : l'ensemble des possibilités est modifié*

Les règles du calcul des probabilités étudiées jusqu'ici sont relatives au cas où l'ensemble Ω des possibilités est donné une fois pour toutes. Dans la pratique de la statistique, il en est rarement ainsi, car certains des événements possibles se réalisent ; l'ensemble des possibilités est dès lors modifié.

Exemple :

Une urne contient dix boules qui sont soit blanches soit noires. On s'intéresse au nombre des boules blanches, mais on ignore ce nombre. L'ensemble des possibilités est :

$$\Omega = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Si maintenant on extrait simultanément 3 boules de l'urne, et si l'on constate que 2 des boules tirées sont blanches, et la troisième noire, il faut exclure les nombres 0, 1 et 10 de l'ensemble des possibilités, puisque le nombre des boules blanches est au moins 2 et au plus 9.

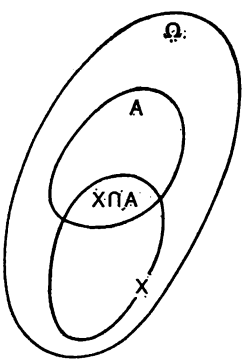
On a un *nouvel* ensemble de possibilités

$$A = \{2, 3, 4, \ldots, 9\}$$

L'ensemble A est une partie de Ω, un *événement* ; on dit que l'événement A est *réalisé* (ici, dans l'expérience du tirage simultané de trois boules).

b) *Tous les événements sont modifiés par la réalisation de l'un d'entre eux*

Si, Ω étant l'ensemble des possibilités, un événement A, partie de Ω, se réalise, on a un nouvel ensemble de possibilités, A ; le simplexe des événements possibles devient alors le simplexe des parties de A, organisées par inclusion ; pour tout événement X de l'ensemble Ω des possibilités (partie de Ω), seule la partie $X \cap A$ commune à A et X reste un événement possible lorsque A est réalisé : les seuls cas (éléments de Ω) restant possibles sont en effet ceux de A.

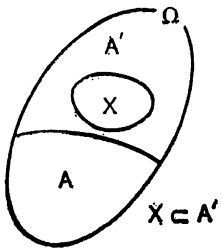


Si A, X devient $X \cap A$

Si A est réalisé, chaque *événement* X est modifié en $X \cap A$.

D'après cette règle de modification des événements tout événement X inclus dans la partie complémentaire A′ de A dans Ω est modifié en :

$$X \cap A = \varnothing$$



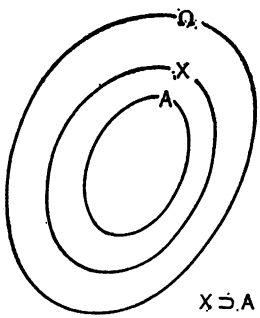
Si A, X devient impossible

Pour le nouvel ensemble A de possibilités, c'est un *événement impossible* ; en effet, la réalisation de A exclut des possibilités l'événement A', et par suite toute partie de A'.

De même, si A est réalisé, tout événement X (partie de Ω) contenant A est d'après la règle modifié en :

$$X \cap A = A$$

Comme A est le nouvel ensemble de possibilités, A est le nouvel *événement certain* ; tout événement contenant A est donc modifié en un événement *certain.*



Si A, X devient certain

En effet, X est réalisé dans tous les cas où A l'est lui-même ; si A se réalise, X se réalise *a fortiori.*

c) *Les probabilités sont modifiées*

La réalisation d'un événement A modifie l'ensemble des possibilités (qui devient A), et modifie chaque événement X (qui devient $X \cap A$) ; lorsque Ω est muni d'une mesure de probabilité, cette réalisation doit modifier non seulement les événements eux-mêmes, mais aussi *les probabilités de ces événements.*

Par exemple, dans l'ensemble Ω de possibilités, un événement X inclus dans A', l'événement A' lui-même, avaient en général une probabilité non nulle ; lorsque A est réalisé, tous ces événements ont pour nouvelle probabilité o puisqu'ils deviennent impossibles ; de même, l'événement A avait en général une probabilité inférieure à 1 ; comme il est devenu *certain*, sa nouvelle probabilité est 1.

On désigne par $p_A(X)$ *la probabilité de l'événement X* si A est réalisé ; $p_A(X)$ se lit « probabilité de X si A » ; on l'appelle souvent *probabilité conditionnelle* de X (sous condition que, ou sachant que, A est réalisé). On dit aussi probabilité *a posteriori* de X ; par opposition, on appelle dans ce cas probabilité *a priori* de X sa probabilité avant que A soit réalisé.

$$p_A(X) : \text{Pr. de X si A est réalisé}$$

On note parfois $p(X \mid A)$ ou $p(X \text{ si } A)$ au lieu de $p_A(X)$.

En conformité avec la règle de modification des événements on pose :

$$p_A(X) = p_A(X \cap A)$$

La probabilité, si A est réalisé, de l'événement X est la probabilité, si A est réalisé, de l'événement $X \cap A$.

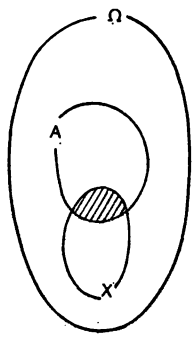
En particulier, on retrouve ainsi que :

Si $X \subset A'$, alors : $p_A(X) = p_A(X \cap A) = p_A(\varnothing) = 0$
Si $X \supset A$, alors : $p_A(X) = p_A(X \cap A) = p_A(A) = 1$

d) *La règle de modification des probabilités : règle de proportionnalité*

Dans les autres cas, c'est-à-dire dans ceux où l'événement X n'est pas inclus dans A' ou ne contient pas A, il faut une règle qui permette de passer de la probabilité de X *avant que* A se réalise (sa probabilité pour l'ensemble Ω de possibilités), notée dans la suite $p_\Omega(X)$, à sa probabilité conditionnelle, $p_A(X)$.

Plus précisément, on avait *une mesure de probabilité* p_Ω sur Ω qui, à chaque événement inclus dans A (mais par rapport à l'ensemble Ω des possibilités), faisait correspondre un nombre, sa probabilité. Lorsque A est réalisé, on a une nouvelle mesure de probabilité p_A qui à chaque partie X de A (événement de l'ensemble A de possibilités) fait correspondre sa probabilité *modifiée* par le fait que A est réalisé, que l'ensemble des possibilités n'est plus que A (1).



(1) p_Ω peut être interprétée comme une expression numérique de *l'opinion* que l'on a sur l'ensemble Ω ; la règle de passage de p_Ω à p_A indique comment modifier cette opinion à la suite de la réalisation de A.

Il est simple de prendre comme règle de modification des mesures de probabilité une règle de *proportionnalité* : par *définition,* la probabilité conditionnelle (si A) d'*un événement* $X \subset \Omega$ *est proportionnelle à la probabilité de* $X \cap A$ *avant que A soit réalisé.* Soit λ le coefficient de proportionnalité :

$$p_A(X) = p_A(X \cap A) = \lambda p_\Omega(X \cap A)$$

Si p_A est définie par cette règle, c'est bien une *mesure de probabilité* ; on va montrer en effet que p_A satisfait aux conditions 1, 2 et 3 du § 2 du chapitre VI, pourvu que le coefficient λ soit choisi convenablement.

— Comment faut-il choisir λ ? Comme en particulier A est une partie de A (sa partie pleine), il faut que :

$$p_A(A) = \lambda p_\Omega(A)$$

Or : $p_A(A) = 1$ (d'après la condition 1) :

$$1 = \lambda p_\Omega(A)$$

Deux cas sont alors possibles :

a) Si $p_\Omega(A) = 0$ (si l'événement A avait une probabilité nulle), alors aucune valeur de λ ne convient, car :

$$\forall \lambda, \qquad \lambda \times 0 = 0 \neq 1$$

La probabilité conditionnelle d'une partie X de A n'est pas définie : *il n'y a pas de mesure de probabilité conditionnelle* (si A est réalisé) *lorsque la probabilité de A est nulle.*

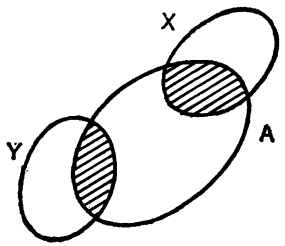
b) Si $p_\Omega(A) > 0$ (lorsque l'événement A avait une probabilité non nulle), alors :

$$\lambda = \frac{1}{p_\Omega(A)} \quad \text{et} \quad p_A(X) = \frac{p_\Omega(X \cap A)}{p_\Omega(A)}$$

Ce choix de λ garantit que les conditions 1 et 2 des mesures de probabilité sont bien satisfaites par la mesure conditionnelle p_A.

1. $p_A(X) \geqslant 0$, puisque $\lambda = \frac{1}{p_\Omega(A)} > 0$.

2. $p_A(A) = 1$, puisque $p_A(A) = \frac{p_\Omega(A)}{p_\Omega(A)} = 1$.



3. Quant à la condition 3 d'additivité, elle se vérifie immédiatement :

— Si $X \cap Y = \varnothing$, comme

$$(X \cup Y) \cap A = (X \cap A) \cup (Y \cap A)$$

(distributivité de $\cap$ pour $\cup$), et $(X \cap A) \cap (Y \cap A) = \varnothing$, on peut écrire, en se servant de l'additivité de p_Ω :

$$p_A(X \cup Y) = \frac{p_\Omega[(X \cup Y) \cap A]}{p_\Omega(A)} = \frac{p_\Omega[(X \cap A) \cup (Y \cap A)]}{p_\Omega(A)}$$

$$= \frac{p_\Omega(X \cap A) + p_\Omega(Y \cap A)}{p_\Omega(A)} = p_A(X) + p_A(Y)$$

On peut donc bien adopter comme règle de modification des probabilités, c'est-à-dire comme *définition* des probabilités conditionnelles :

Si $p_\Omega(A) > 0$, pour toute partie X de Ω :

$$p_A(X) = \frac{p_\Omega(X \cap A)}{p_\Omega(A)}$$

Cette règle s'énonce :

La probabilité conditionnelle, si A est réalisé, d'un événement X, est le rapport de la probabilité de l'intersection de X et A à la probabilité de A, supposée non nulle.

Il ne faut pas oublier l'énoncé complémentaire :

Si $p_\Omega(A) = 0$, la mesure de probabilité conditionnelle (si A est réalisé) *n'est pas définie.*

Exemple :

Soit $\Omega = \{1, 2, 3\}$, et p_Ω définie par la distribution :

$$p_1 = 1/6 \qquad p_2 = 1/3 \qquad p_3 = 1/2$$

On peut alors dresser un tableau de *toutes* les mesures de probabilité conditionnelle, en prenant successivement pour A chacune de 7 parties de probabilité non nulle de Ω : toutes les parties non vides, puisque dans cet exemple seule la partie $\varnothing$ a une probabilité nulle.

Sur le tableau, on lit en colonne les probabilités conditionnelles ; à chaque colonne correspond un événement considéré comme réalisé.

La première colonne, dans laquelle c'est Ω qui est réalisé, donne bien entendu la mesure p_Ω elle-même.

Tableau des $p_A(X)$

X \ A	Ω	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1\}$	$\{2\}$	$\{3\}$
Ω	1	1	1	1	1	1	1
$\{1,2\}$	1/2	1	1/4	2/5	1	1	0
$\{1,3\}$	2/3	1/3	1	3/5	1	0	1
$\{2,3\}$	5/6	2/3	3/4	1	0	1	1
$\{1\}$	1/6	1/3	1/4	0	1	0	0
$\{2\}$	1/3	2/3	0	2/5	0	1	0
$\{3\}$	1/2	0	3/4	3/5	0	0	1
$\varnothing$	0	0	0	0	0	0	0

Chacune des probabilités se calcule de la façon indiquée ci-dessous pour deux exemples :

— Si $A=\{2,3\}$ et $X=\{1,2\}$, alors $A\cap X=\{2\}$:

$$p_{\{2,3\}}(\{1,2\})=\frac{p_\Omega(\{1,2\}\cap\{2,3\})}{p_\Omega(\{2,3\})}=\frac{p_\Omega(\{2\})}{p_\Omega(\{2,3\})}$$

$$=\frac{p_2}{p_2+p_3}=\frac{1/3}{1/3+1/2}=\frac{2}{5}$$

— Si $A=\{1\}$ et $X=\{1,3\}$, alors $A\cap X=\{1\}$:

$$p_{\{1\}}(\{1,3\})=\frac{p_\Omega\{1\}}{p_\Omega\{1\}}=1 \qquad \text{(ici, } X\supset A\text{)}$$

2. Produit cartésien et croisement d'aléas

a) *Produit cartésien et croisement de partitions*

Soit une population, assimilée à un ensemble Ω ; à chaque individu de cette population on fait correspondre d'une part sa classe d'âge, d'autre part sa tranche de revenu. Supposons pour fixer les idées que l'on se soit donné 4 classes d'âge, désignées par y_1, y_2, y_3, y_4 :

y_1 : < 25 ans ; y_2 : 25-34 ans

y_3 : 35-44 ans ; y_4 : $\geqslant 45$ ans

De même, supposons qu'il y ait trois tranches de revenu z_1, z_2, z_3 :

z_1 : $<$ 500 F par mois ; z_2 : 500 — 1 000 ; z_3 : $>$ 1 000 F

On a ainsi deux applications $f : \Omega \to \Omega'_1$ et $g : \Omega \to \Omega'_2$ où l'on a posé :

$$\Omega'_1 = \{ y_1, y_2, y_3, y_4 \}$$
$$\Omega'_2 = \{ z_1, z_2, z_3 \}$$

Chacune de ces applications définit sur Ω un quotient, Ω/f et Ω/g respectivement ; l'ensemble quotient par f comprend (en général, si f est surjective) quatre classes ; on le désigne par Ω_1 :

$$\Omega_1 = \{ A_1, A_2, A_3, A_4 \}$$

La classe A_i est constituée des éléments de Ω qui ont y^i pour image par f.

De même, à g correspond le quotient :

$$\Omega_2 = \{ B_1, B_2, B_3 \}$$

Si à chaque élément de Ω on fait correspondre à la fois son âge et son revenu, on définit une application h de Ω dans le produit cartésien $\Omega'_1 \times \Omega'_2$; à chaque élément de Ω correspond le *couple* de ses images par f et par g :

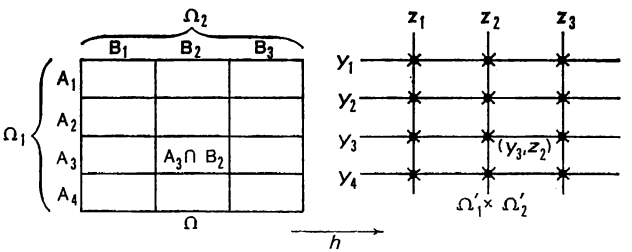
$$h : \Omega \to \Omega'_1 \times \Omega'_2$$
$$\forall x \in \Omega \qquad hx = (fx, gx)$$

L'application h induit dans Ω une partition, la *partition croisée* de Ω/f et Ω/g, dans laquelle deux éléments de Ω appartiennent à la même classe si et seulement si ils ont même image par f et par g (cf. *MSH*, chap. X, § 5). Les classes du quotient Ω/h sont donc les intersections deux à deux des classes A_i et des classes B_j ; dans l'exemple, il y en a au plus $4 \times 3 = 12$ (certaines peuvent éventuellement être vides) ; dans la classe $A_3 \cap B_2$ par exemple se trouvent les éléments x tels que :

$$hx = (y_3, z_2)$$

i.e. $fx = y_3$ et $gx = z_2$.

i.e. la tranche d'âge de x est 35-44, sa tranche de revenu 500-1 000.



L'application $h : \Omega \to \Omega'_1 \times \Omega'_2$ peut se noter (f, g) ; h est en effet définie par le *couple* des applications f et g.

b) *Produit cartésien et croisement d'aléas*

On suppose maintenant que Ω est muni d'une mesure de probabilité p : on a un aléa (Ω, p). C'est le cas, par exemple, si l'on extrait « au hasard » un individu de Ω ; si les précautions expérimentales nécessaires sont prises, cela peut signifier que chaque individu a une même probabilité d'être extrait ; p est alors la mesure uniforme sur Ω.

A l'application $f : \Omega \to \Omega'_1$ correspondent deux aléas, répliques l'un de l'autre (chap. VII, § 7) :

1° Son *aléa image* $(\Omega'_1, f\mathbf{p})$:

Ω'_1	y_1	y_2	y_3	y_4
$f\mathbf{p}$	$p(A_1)$	$p(A_2)$	$p(A_3)$	$p(A_4)$

A y_i est attribuée la probabilité $p(A_i)$ de la classe A_i dont les éléments, ont y_i pour image par f.

2° Son *aléa quotient* $(\Omega_1, f\mathbf{p})$:

Ω_1	A_1	A_2	A_3	A_4
$f\mathbf{p}$	$p(A_1)$	$p(A_2)$	$p(A_3)$	$p(A_4)$

Si en particulier, p est uniforme, on a :

$$p(A_1) = \frac{|A|}{|\Omega|}, \quad \ldots, \quad p(A_4) = \frac{|A_4|}{|\Omega|}$$

De même, à l'application $g : \Omega \to \Omega'_2$ correspondent les deux aléas :

1° Image

Ω'_2	z_1	z_2	z_3
$g\mathbf{p}$	$p(B_1)$	$p(B_2)$	$p(B_3)$

2° Quotient

Ω_2	B_1	B_2	B_3
g	$p(B_1)$	$p(B_2)$	$p(B_3)$

Les aléas image et quotient associés à f donnent la distribution de probabilité des *âges* dans la population ; ceux qui sont associés à g donnent la distribution des revenus. $p(A_1)$ est la probabilité pour que l'individu tiré ait moins de 25 ans ; $p(B_2)$ est la probabilité pour qu'il ait un revenu compris entre 500 et 1 000 F par mois.

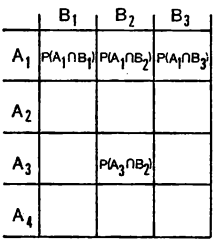
A l'application $h : \Omega \to \Omega'_1 \times \Omega'_2$ correspondent enfin deux aléas :

1° Un *aléa image*, ayant pour ensemble de définition le produit cartésien $\Omega'_1 \times \Omega'_2$, et dans lequel chaque couple (y_i, z_j) a pour probabilité q_{ij} celle de la classe $A_i \cap B_j$ correspondante :

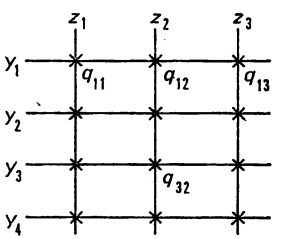
$$q_{ij} = p(A_i \cap B_j)$$

2° Un *aléa quotient*, ayant pour ensemble de définition l'ensemble des 12 classes $A_i \cap B_j$, chacune étant affectée de sa probabilité $p(A_i \cap B_j)$ selon p.

Ces deux aléas sont répliques l'un de l'autre.



$(\Omega/h, q)$



$(\Omega'_1 \times \Omega'_2, q)$

$$q = hp$$

Ces aléas fournissent la distribution de probabilité du couple âge-revenu ; $q_{32} = p\,(A_3 \cap B_2)$ est la probabilité pour que l'individu tiré soit dans la troisième tranche d'âge et la seconde tranche de revenu.

On peut dire que l'aléa $(\Omega'_1 \times \Omega'_2, q)$ est le *produit cartésien* des aléas (Ω'_1, fp) et (Ω'_2, gp) ; de même, que l'aléa $(\Omega/h, q)$ est le *croisement* des deux aléas $(\Omega/f, fp)$ et $(\Omega/g, gp)$ (1).

c) *Distributions de probabilité conditionnelle associées*

Il peut arriver que l'enquêteur (qui extrait un individu de la population prise en exemple) tombe sur quelqu'un qui lui dise son âge (y_2 par exemple), mais refuse de lui avouer son revenu. Ce dont dispose alors l'enquêteur comme renseignement sur le revenu de l'individu tiré, c'est de la *distribution de probabilité conditionnelle si* A_3, c'est-à-dire de la distribution conditionnelle du revenu, connaissant l'âge :

$$p_{A_3} = (p_{A_3}(B_1), p_{A_3}(B_2), p_{A_3}(B_3))$$

Sur le tableau croisé des partitions Ω_1 et Ω_2, cette distribution s'obtient *en divisant chaque probabilité de la ligne* A_3 *par la probabilité* p (A_3) *de l'événement* A_3.

Elle peut également se lire sur le tableau $\Omega'_1 \times \Omega'_2$ en divisant les probabilités de couples situés sur la ligne y_3 par la probabilité de y_3, notée $q\,(y_3)$:

$$q\,(y_3) = p\,(A_3)$$

Par ce procédé, à chaque *ligne* du produit est associée une distribution de probabilité conditionnelle : à la ligne A_i (resp. y_i), la distribution *si* A_i (resp. *si* y_i) est réalisé ; elle

(1) Dans la suite de ce chapitre, on dira « produit » pour « produit cartésien ». On ne risque pas, en effet, de confusion avec le produit d'aléas numériques défini au chap. IX.

s'obtient en divisant toutes les probabilités de cette ligne par la probabilité de A_i ; celle-ci est souvent appelée *probabilité marginale* (parce que fréquemment indiquée en marge du tableau).

De même, à chaque colonne B_j, est associée une distribution de probabilité conditionnelle, p_{B_j}. Dans la seconde colonne, par exemple, on a, en divisant toutes les probabilités par $p(B_2)$, la distribution conditionnelle si B_2, p_{B_2} :

$$p_{B_2} = \left(\frac{p(A_1 \cap B_2)}{p(B_2)}, \frac{p(A_2 \cap B_2)}{p(B_2)}, \frac{p(A_3 \cap B_2)}{p(B_2)}, \frac{p(A_4 \cap B_2)}{p(B_2)} \right)$$

Exemple numérique :

On suppose que les probabilités du tableau ont les valeurs indiquées dans les cases (vérifiez que la somme de ces valeurs vaut bien 1) :

	B_1	B_2	B_3
A_1	0,05	0,10	0,20
A_2	0,02	0,15	0,03
A_3	0,03	0,09	0,03
A_4	0,20	0,06	0,04

Les probabilités *marginales* correspondantes, qui sont les probabilités des événements A_i (en lignes) et B_j (en colonnes) s'obtiennent en additionnant toutes les probabilités d'une ligne (resp. d'une colonne).

En effet, A_1 par exemple est partitionné par B_1, B_2 et B_3 en : $A_1 \cap B_1$, $A_1 \cap B_2$, $A_1 \cap B_3$.

On a donc :

$$\begin{aligned} p(A_1) &= p(A_1 \cap B_1) + p(A_1 \cap B_2) + p(A_1 \cap B_3) \\ &= 0{,}05 + 0{,}10 + 0{,}20 = 0{,}35 \end{aligned}$$

D'où les probabilités marginales :

	$p(B_1)$ 0,30	$p(B_2)$ 0,40	$p(B_3)$ 0,30
$p(A_1) = 0{,}35$			
$p(A_2) = 0{,}20$			
$p(A_3) = 0{,}15$			
$p(A_4) = 0{,}30$			

Les deux séries de distributions de probabilité conditionnelles associées sont alors :

Lignes

	B_1	B_2	B_3
p_{A_1}	5/35	10/35	20/35
p_{A_2}	2/20	15/20	3/20
p_{A_3}	3/15	9/15	3/15
p_{A_4}	20/30	6/30	4/30

Colonnes

	p_{B_1}	p_{B_2}	p_{B_3}
A_1	5/30	10/40	20/30
A_2	2/30	15/40	3/30
A_3	3/30	9/40	3/30
A_4	20/30	6/40	4/30

Bien entendu, tous ces calculs auraient pu être faits sur le produit cartésien $\Omega'_1 \times \Omega'_2$.

On remarque que pour *les distributions conditionnelles en « lignes », la somme des termes d'une ligne vaut 1*, comme il se doit ; de même pour les distributions conditionnelles en « colonnes », la somme des nombres d'une colonne vaut 1. Vérifiez-le (1).

(1) En inférence statistique, produits et croisements d'aléas interviennent en particulier dans la théorie des *tableaux de dépendance* (angl. : tableaux de contingence).

Remarque :

Il est souvent commode de figurer les produits et croisements d'aléas par des arbres, plutôt que par des tableaux. Cf. chap. IX, § 2-*e*.

d) *Formule de Bayes*

Considérons de nouveau l'événement $A_3 \cap B_2$, qui correspond au couple (y_3, z_2). Sa probabilité $p(A_3 \cap B_2)$ fournit deux probabilités conditionnelles, l'une en ligne :

$$p_{A_3}(B_2) = \frac{p(A_3 \cap B_2)}{p(A_3)} \qquad \text{(dans l'exemple numérique : 9/15)}$$

l'autre en colonne :

$$p_{B_2}(A_3) = \frac{p(A_3 \cap B_2)}{p(B_2)} \qquad \text{(dans l'exemple numérique : 9/40)}$$

Ces deux égalités s'écrivent encore :

$$p(A_3)\, p_{A_3}(B_2) = p(A_3 \cap B_2) = p(B_2)\, p_{B_2}(A_3)$$

Elles permettent de relier la probabilité conditionnelle *si* B_2 aux probabilités marginales, et aux probabilités conditionnelles *si* A_i.

$$p_{B_2}(A_3) = \frac{p(A_3)\, p_{A_3}(B_2)}{p(B_2)}$$

De même, on calcule :

$$p_{B_2}(A_1) = \frac{p(A_1)\, p_{A_1}(B_2)}{p(B_2)}$$

$$p_{B_2}(A_2) = \frac{p(A_2)\, p_{A_2}(B_2)}{p(B_2)}$$

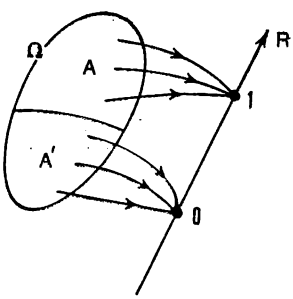
etc.

Ces formules, qui permettent de calculer chaque distribution conditionnelle *si* B_j (en colonne) à partir des distributions conditionnelles *si* A_i (en lignes) et des distributions *marginales*, sont connues sous le nom de *formules de Bayes*.

e) *Produit cartésien d'aléas de Bernoulli*

Dans la pratique de la statistique ce sont très souvent deux aléas de Bernoulli, images par deux applications f et g d'un même aléa (Ω, p), que l'on croise (ou dont on fait le produit). C'est le cas, par exemple, chaque fois que l'on considère deux caractères α et β dichotomiques dans une population dont on tire un individu. Les tableaux se simplifient dans ce cas particulier.

L'application f qui fait correspondre à chaque $x \in \Omega$ le nombre 1 si x présente le caractère α, et 0 sinon, a pour image et pour quotient :



0	1
$(1 - p(A))$	$p(A)$

Aléa image

A'	A
$(1 - p(A))$	$p(A)$

Aléa quotient

A est la partie de Ω constituée des individus présentant le caractère α.

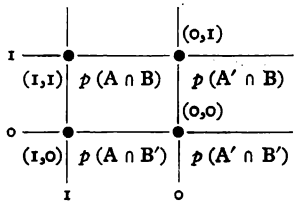
De même, au caractère β est associée une application g, et une partie B de Ω ; un aléa image et un aléa quotient.

Le croisement et le produit des aléas définis par f et g

respectivement sont donnés par les tableaux 2×2 ci-dessous :

		A	A′
		p (A)	p (A′)
B	p (B)	p (A ∩ B)	p (A′ ∩ B)
B′	p (B′)	p (A ∩ B′)	p (A′ ∩ B′)

Croisement
des aléas quotients



Produit
des aléas images

3. Aléas et événements indépendants en probabilité

a) *Aléas indépendants*

Soit deux aléas *images d'un même aléa* (Ω, p) par deux applications f et g de Ω dans Ω'_1 et Ω'_2 respectivement : $(\Omega'_1, \mathbf{q}^{(1)})$ et $(\Omega'_2, \mathbf{q}^{(2)})$ où $\mathbf{q}^{(1)} = f\mathbf{p}$ est la distribution image de $\mathbf{p}$ par f et $\mathbf{q}^{(2)} = g\mathbf{p}$ est la distribution image de $\mathbf{p}$ par g.

Ces deux aléas sont dits *indépendants en probabilité par rapport à* p si et seulement si *la distribution de leur aléa produit* s'obtient par la règle (les notations sont celles employées au § 2) :

$$\forall\, (i,j) \in \Omega'_2 \times \Omega'_1 : q_{ij} = q_i^{(1)}\, q_j^{(2)}$$

Il est équivalent de dire que les deux aléas sont indépendants en probabilité par rapport à p si et seulement si la distribution de probabilité du croisement des aléas quotients de (Ω, p) par f et g s'obtient par la règle :

$$\forall_i, \forall_j : \boxed{p\,(\mathrm{A}_i \cap \mathrm{B}_j) = p\,(\mathrm{A}_i) \times p\,(\mathrm{B}_j)}$$

Autre façon de dire, la distribution sur le produit (le croisement) s'obtient en *multipliant ligne par colonne les probabilités marginales.*

Exemple :

		A_1	A_2	A_3
		1/3	1/6	1/2
B_1	1/4	1/12	1/24	1/8
B_2	3/4	1/4	1/8	3/8

On a :

$$p(A_1 \cap B_1) = p(A_1)\, p(B_1) = 1/3 \times 1/4 = 1/12 \text{ etc.}$$

Contre-exemple :

Les deux aléas qui ont servi d'exemple au § 2 ne sont pas indépendants ; on a en effet pour ces aléas :

$p(A_1 \cap B_1) = 0{,}05 \quad p(A_1)\, p(B_1) = 0{,}30 \times 0{,}35 = 0{,}105$

etc.

Remarque :

Si deux aléas images d'un même aléa sont indépendants en probabilité, il est clair que les distributions conditionnelles si A_i sont identiques aux distributions *marginales* ; on a en effet dans ce cas :

$$p_{A_i}(B_j) = \frac{p(A_i \cap B_j)}{p(A_i)} = \frac{p(A_i)\, p(B_j)}{p(A_i)} = p(B_j)$$

De même pour les distributions conditionnelles si B_j :

$$p_{B_j}(A_i) = p(A_i)$$

C'est en ce sens que l'on dit souvent que si deux aléas sont indépendants la connaissance de la valeur prise par l'un n'apporte aucune information supplémentaire sur l'autre ; si l'on sait que l'événement A_i est réalisé, l'événement B_j a même probabilité $p(B_j)$ que si l'on ne sait rien sur la réalisation de A_i.

b) *Événements indépendants*

Deux événements A et B d'un même ensemble Ω probabilisé par une mesure p sont dits indépendants *par rapport à* p si les deux aléas quotients de (Ω, p) par les partitions $\{A, A'\}$ définie par A d'une part, $\{B, B'\}$ définie par B d'autre part, sont indépendants. C'est-à-dire que le tableau du croisement de ces partitions est :

		A	A′
		$p(A)$	$p(A')$
B	$p(B)$	$p(A)\,p(B)$	$p(B)\,p(A')$
B′	$p(B')$	$p(A)\,p(B')$	$p(A')\,p(B')$

$$p(A \cap B) = p(A)\,p(B)$$
$$p(A \cap B') = p(A)\,p(B')$$
$$p(A' \cap B) = p(A')\,p(B)$$
$$p(A' \cap B') = p(A')\,p(B')$$

Exercice :

— Montrer que si $p(A)$ et $p(B)$ ne valent ni zéro, ni un, c'est-à-dire si A et B ne sont ni certains, ni impossibles :

$$0 < p(A) < 1 \qquad \text{et} \qquad 0 < p(B) < 1$$

il faut et suffit pour que les événements A et B soient indépendants par rapport à p que :

$$p(A \cap B) = p(A)\,p(B)$$

Indication : les trois autres égalités qu'implique la définition sont des conséquences de la première.

Remarque :

La notion d'indépendance en probabilité d'aléas ou d'événements n'a de sens que *par rapport à une mesure de probabilité* p. Pour le même ensemble Ω, deux parties A et B (deux événements), qui sont indépendants pour une certaine probabilisation de Ω, ne le sont pas pour une autre.

Il est fréquent en inférence statistique que l'on fasse sur deux aléas l'hypothèse de leur indépendance pour déterminer la distribution de probabilité de leur croisement (ou de leur produit) à partir des distributions marginales supposées connues *a priori*. On va en voir un exemple.

c) *Application : distribution binomiale*

On tire d'une « urne » ayant une proportion θ de boules « blanches » (et $(1 - \theta)$ de boules « non blanches ») n boules successivement, avec remise.

L'ensemble Ω des possibilités est, on l'a vu au chapitre VII, § 5 et 6, l'ensemble des 2^n suites de n symboles B (blanche) ou N (non blanche).

A chaque tirage, on peut faire correspondre un aléa de Bernoulli (cf. chap. VII, § 3) :

0	1
$(1 - \theta)$	θ

Cet aléa prend la valeur 0, avec la probabilité $1 - \theta$, si la boule tirée est non blanche ; et il prend la valeur 1 sinon.

Si l'on procède chaque fois à la remise de la boule tirée dans l'urne, on peut (moyennant des précautions expérimentales) faire l'*hypothèse* que les aléas de Bernoulli relatifs à deux tirages distincts sont indépendants en probabilité ; moyennant cette hypothèse, les deux premiers tirages fournissent le tableau suivant.

		Première boule : B, θ	Première boule : N, $1-\theta$
Deuxième boule	B, θ	θ^2	$\theta(1-\theta)$
	N, $1-\theta$	$(1-\theta)\theta$	$(1-\theta)^2$

Le fait que l'on attribue la probabilité $\theta(1-\theta)$ par exemple à l'événement « la première boule tirée est N et la seconde est B », c'est-à-dire à la partie de Ω constituée par toutes les suites commençant par NB..., résulte de l'*hypothèse d'indépendance.*

Dans la même hypothèse, l'aléa défini par la troisième boule tirée est indépendant de celui qui est défini par le couple des deux premiers tirages. On a pour les trois premiers tirages un quotient, donné par le tableau suivant :

La suite des deux premières boules tirées	Troisième boule : B, θ	Troisième boule : N, $1-\theta$
BB, θ^2	θ^3	$(1-\theta)\theta^2$
BN, $\theta(1-\theta)$	$\theta^2(1-\theta)$	$\theta(1-\theta)^2$
NB, $(1-\theta)\theta$	$(1-\theta)\theta^2$	$(1-\theta)^2\theta$
NN, $(1-\theta)^2$	$(1-\theta)^2\theta$	$(1-\theta)^3$

L'aléa produit des trois aléas de Bernoulli correspondant aux trois premiers tirages a ainsi la distribution de probabilité donnée par le tableau ci-dessous ; à ce tableau, on a ajouté l'indication du nombre de boules B et du nombre de boules N de chaque suite (1).

		Nombre de B	Nombre de N
BBB	θ^3	3	0
BBN	$\theta^2 (1 - \theta)$	2	1
BNB	$\theta^2 (1 - \theta)$	2	1
BNN	$\theta (1 - \theta)^2$	1	2
NBB	$\theta^2 (1 - \theta)$	2	1
NBN	$\theta (1 - \theta)^2$	1	2
NNB	$\theta (1 - \theta)^2$	1	2
NNN	$(1 - \theta)^3$	0	3

On constate que chaque suite de trois symboles B et N a une probabilité égale à θ élevé à la puissance « nombre de B » multiplié par $(1 - \theta)$ élevé à la puissance « nombre de N ».

Ceci est général. Si une suite de n tirages a fourni k boules blanches et $(n - k)$ boules non blanches, sa probabilité est :

$$\theta^k (1 - \theta)^{n-k}$$

En effet les n aléas de Bernoulli relatifs au premier tirage, au second tirage, etc., sont n aléas *indépendants en probabilité* ; à chaque tirage supplémentaire, une suite donnée peut se prolonger soit par B, et sa probabilité est alors multipliée par θ, soit par N, et sa probabilité est alors multipliée par $1 - \theta$.

Pour obtenir l'aléa binomial il faut ensuite faire correspondre à chacune de 2^n suites le nombre des B qu'elle comporte ; ce nombre est encore appelé la *fréquence* de B

(1) Reprenez ce paragraphe en remplaçant les tableaux par des schémas en arbres, comme c'est indiqué chap. IX, § 2-*e*.

dans la suite des tirages. La probabilité pour que cet aléa binomial prenne la valeur k, c'est donc la probabilité de la partie de Ω (ensemble des 2^n suites possibles) constituée par les $\binom{n}{k}$ suites comportant k fois la lettre B. Comme celles-ci ont toutes même probabilité $\theta^k (1-\theta)^{n-k}$ la probabilité cherchée est :

$$b_k = \binom{n}{k} \theta^k (1-\theta)^{n-k}$$

En particulier, la probabilité de la suite ne comportant que des B est :

$$b_n = \binom{n}{n} \theta^n (1-\theta)^0 = 1 \times \theta^n \times 1 = \theta^n$$

La probabilité de la suite ne comportant que des N est :

$$b_0 = \binom{n}{0} \theta^0 (1-\theta)^n = (1-\theta)^n$$

Exercice :

Déduire de ce qui précède que si a et b sont deux nombres positifs, de somme égale à 1 ($a + b = 1$), on a :

$$1 = \binom{n}{0} a^n + \binom{n}{1} a^{n-1} b + \ldots + \binom{n}{k} a^{n-k} b^k + \ldots + \binom{n}{n} b^n$$

— En déduire que si u et v sont deux nombres positifs quelconques tels que $u + v \neq 0$, on a :

$$(u+v)^n = \binom{n}{0} u^n + \binom{n}{1} u^{n-1} v + \ldots + \binom{n}{n} v^n$$

(formule dite du « binôme de Newton »).

Récapitulation

Ω est l'ensemble des possibilités ; un événement $A \subset \Omega$ se réalise ($p(A) > 0$).

Tout événement X $\xrightarrow[\text{en}]{\text{est modifié}}$ $X \cap A$ (X si A)

Sa probabilité $p(X)$ $\xrightarrow[\text{en}]{\text{est modifiée}}$ $\dfrac{p(X \cap A)}{p(A)}$ (Pr. de X si A)

On note : $p_A(X) = \dfrac{p(X \cap A)}{p(A)}$

Croisement (produit cartésien) d'aléas

1. Deux partitions $A_1, A_2, \ldots$; $B_1, B_2, \ldots$; de l'ensemble Ω des possibilités.

2. Le croisement des partitions :

	A_1	A_2	...
B_1			
B_2			
⋮			

3. Pour chaque classe de la partition croisée : sa probabilité $p\,(A_i \cap B_j)$.

4. Pour chaque ligne (colonne) du tableau, une probabilité *marginale*, somme des probabilités de la ligne (colonne) :

$$p\,(A_i) = \sum_j p\,(A_i \cap B_j)$$

$$p\,(B_j) = \sum_i p\,(A_i \cap B_j)$$

	$p\,(A_1)$	$p\,(A_2)$	...
$p\,(B_1)$	$p\,(A_1 \cap B_1)$	$p\,(A_2 \cap B_1)$	...
$p\,(B_2)$	$p\,(B_2 \cap A_1)$	$p\,(A_2 \cap B_2)$	...
⋮	⋮	⋮	

5. Dans chaque ligne (colonne) une distribution *conditionnelle* si B_j (si A_i) : on divise la ligne (colonne) par sa probabilité marginale :

$$p_{B_j}(A_i) = \frac{p\,(A_i \cap B_j)}{p\,(B_j)} \qquad p_{A_i}(B_j) = \frac{p\,(A_i \cap B_j)}{p\,(A_i)}$$

	A_1	A_2	...
p_{B_1}	$p_{B_1}(A_1)$	$p_{B_1}(A_2)$	...
p_{B_2}	$p_{B_2}(A_1)$	...	...
⋮	⋮	⋮	

	p_{A_1}	p_{A_2}	...
B_1	$p_{A_1}(B_1)$	$p_{A_2}(B_1)$	...
B_2	$p_{A_1}(B_2)$	...	...
⋮	⋮	⋮	

Indépendance en probabilité

Deux aléas images par f et g (deux applications de départ Ω) :
f induit une partition $A_1, A_2, \ldots$, de Ω ;
g induit une partition $B_1, B_2, \ldots$, de Ω.

Ils sont indépendants en probabilité (par rapport à la mesure p) si et seulement si :

$$\forall_{i,j} \qquad p(A_i \cap B_j) = p(A_i)\, p(B_j)$$

En particulier : *Deux événements* A et B ni certains, ni impossibles, sont *indépendants* en probabilité (par rapport à la mesure p) si et seulement si :

$$p(A \cap B) = p(A)\, p(B)$$

CHAPITRE IX

ALÉAS NUMÉRIQUES

1. LES ALÉAS NUMÉRIQUES ET L'ALGÈBRE LINÉAIRE

a) *Aléas numériques*

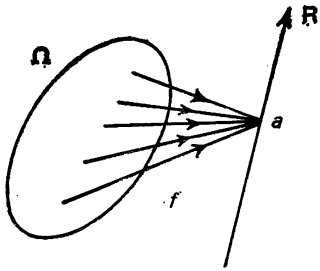
Un aléa (Ω, p) est dit *numérique* si son ensemble Ω de possibilités (ou de définition) est un ensemble de nombres. Les aléas numériques que l'on considère ici sont tous *finis* : leur support est fini (chap. VII, § 1). Selon la nature de l'ensemble Ω (qui peut être **N**, **Z**, **D**, **Q** ou **R**) l'aléa (Ω, p) est dit : aléa entier naturel, aléa entier, aléa décimal, aléa rationnel ou aléa réel.

Par exemple, l'aléa de Bernoulli, l'aléa binomial, l'aléa hypergéométrique (chap. VII, § 3 à 6) sont des aléas entiers naturels ; pour chacun d'entre eux, le support est une partie de **N** donc *a fortiori* de **Q** ou de **R**.

Pour la commodité algébrique (pour pouvoir calculer par addition, soustraction, multiplication et division sur les nombres et faire de l'*algèbre linéaire*), on considère ici désormais que les aléas numériques étudiés sont *rationnels* ou *réels*.

b) *Aléas numériques certains*

Un aléa numérique est dit certain lorsqu'il n'a qu'*une seule valeur possible*.



En particulier, si (Ω, p) est un aléa quelconque, et f une *application constante*, d'image a, de Ω dans $\mathbf{R}$:

$$\forall x \in \Omega,\ fx = a$$

alors l'aléa numérique *image* de l'aléa (Ω, p) par f est un aléa certain, de valeur a.

c) *Somme et homothétiques d'aléas numériques images d'un même aléa*

— Si (Ω, p) est un aléa, numérique ou non, *donné une fois pour toutes*, à chaque application $f : \Omega \to \mathbf{R}$ (ou $\mathbf{Q}$) correspond un aléa numérique *image* $(f\Omega, fp)$ de l'aléa (Ω, p) par f (chap. VII, § 2).

L'ensemble $\mathbf{R}^{\Omega}$ de *toutes* les applications de Ω dans $\mathbf{R}$ ($\mathbf{Q}^{\Omega}$, si l'ensemble d'arrivée est $\mathbf{Q}$) peut être muni d'une *structure de vectoriel* (chap. III, § 1 et *MSH*, chap. XI) en posant, par définition de la somme de deux applications f et g :

$$f + g = h \Leftrightarrow \forall x \in \Omega,\ fx + gx = hx$$

et par définition du produit d'une application par un nombre (scalaire) λ (*homothétie* de rapport λ) :

$$\lambda f = h \Leftrightarrow \forall x \in \Omega,\ \lambda f = hx$$

L'aléa numérique image de (Ω, p) par l'application $f + g = h$ de Ω dans $\mathbf{R}$ (ou $\mathbf{Q}$) est appelé *somme* des deux aléas images de (Ω, p) par f et g respectivement.

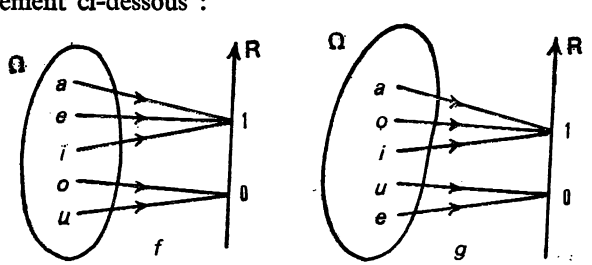
De même :

L'aléa numérique image de (Ω, p) par l'application λf de Ω dans $\mathbf{R}$ (ou $\mathbf{Q}$) est appelé *homothétique* dans le rapport λ *(ou produit par* λ*)* de l'aléa image de (Ω, p) par f.

Exemple :

$\Omega = \{ a, e, i, o, u \}$ $\quad$ $\mathbf{p} = (1/10, 4/10, 3/10, 1/10, 1/10)$

f et g sont les applications de Ω dans $\mathbf{R}$ données graphiquement ci-dessous :

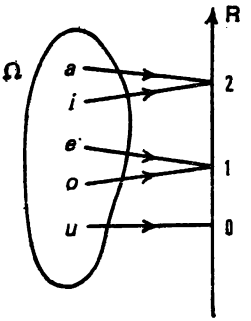


Les aléas images par f et g sont les deux aléas de Bernoulli :

$f\Omega$	0	1
$f\mathbf{p}$	2/10	8/10

$g\Omega$	0	1
$g\mathbf{p}$	5/10	5/10

La somme $f + g = h$:



		f		
		1 $\{a, e, i\}$	0 $\{o, u\}$	Pr. ↓
g	1 — a o i	$\{a, i\}$ $1 + 1 = 2$	$\{o\}$ $1 + 0 = 1$	5/10
	0 — e u	$\{e\}$ $0 + 1 = 1$	$\{u\}$ $0 + 0 = 0$	5/10
	Pr. →	8/10	2/10	

La somme a pour support $h\Omega = \{0, 1, 2\}$.

La distribution de probabilité de la somme est l'image par h de $\mathbf{p}$, soit :

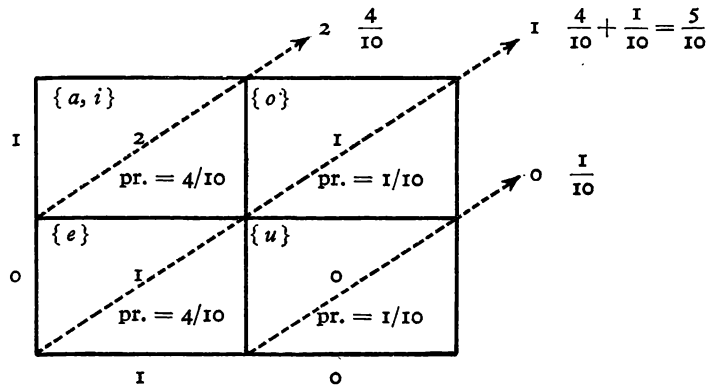
$$h\mathbf{p} = \mathbf{p}' = (p'_0, p'_1, p'_2)$$

$$p'_0 = p(\{u\}) = 1/10$$

$$p'_1 = p(\{e, o\}) = 4/10 + 1/10 = 5/10$$

$$p'_2 = p(\{a, i\}) = 1/10 + 3/10 = 4/10$$

Sur le tableau croisant les partitions induites sur Ω par f et g respectivement, on peut lire directement cette somme, en additionnant en diagonale les probabilités des cases :



L'homothétique dans le rapport 3/2 de l'aléa image de (Ω, p) par f est l'aléa dont le support est $\{ 0, 3/2 \}$, car :

$$3/2 \times 0 = 0 \qquad 3/2 \times 1 = 3/2$$

Sa distribution de probabilité est évidemment donnée par :

	0	3/2
Pr. :	2/10	8/10

d) *Vectoriel des aléas numériques image d'un même aléa* (Ω, p)

— L'ensemble des aléas numériques images, dans un même *corps* de nombres (**R** ou **Q**), d'un *même* aléa (Ω, p), est donc tel que l'on puisse les additionner deux à deux et

définir des homothétiques. Ils constituent un *vectoriel* : le vectoriel des aléas numériques images d'un même aléa (Ω, p).

Comme (Ω, p) est donné *une fois pour toutes*, et qu'il y a par conséquent correspondance biunivoque entre les applications f de Ω dans $\mathbf{R}$, et les aléas numériques images $(f\Omega, fp)$, on désigne dans la suite, par abus de langage, mais pour alléger celui-ci, chaque aléa numérique par l'application f qui le définit ; la phrase : « l'aléa numérique f » signifiera en fait « l'aléa numérique image de (Ω, p) par l'application $f : \Omega \to \mathbf{R}$ ». Cette convention n'a de sens que grâce à la donnée une fois pour toutes de (Ω, p) ; et le calcul fait ci-dessus de la distribution de probabilité d'une somme de deux aléas numériques montre bien que ce calcul n'est possible que grâce à la connaissance de l'ensemble Ω et de sa mesure de probabilité p.

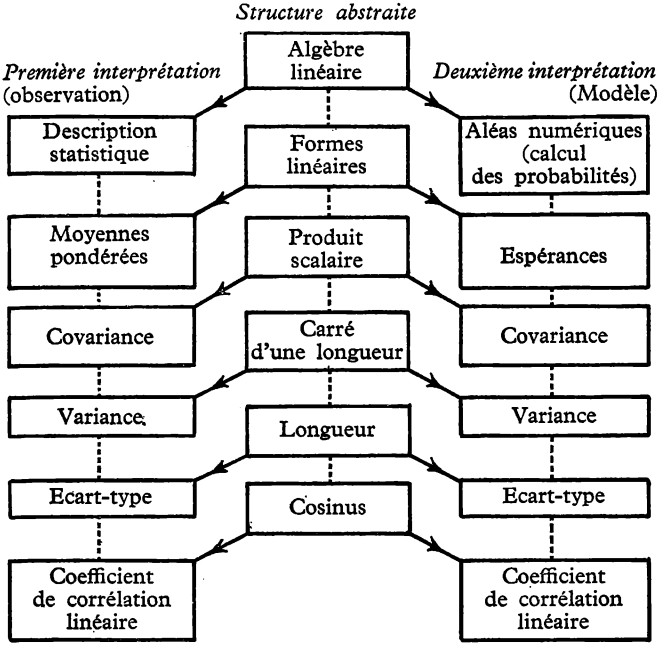
— Dans le vectoriel des aléas numériques images d'un même aléa numérique, les concepts de l'algèbre linéaire tels que *forme linéaire*, *produit scalaire*, *cosinus*, *longueur*, vont avoir leur interprétation en calcul des probabilités, de même qu'ils avaient une interprétation en description statistique, à savoir respectivement : *moyenne* (pondérée), *covariance*, *coefficient de corrélation linéaire*, *écart type* (cf. chap. III, IV, V).

En ce qui concerne les aléas numériques, sauf pour les formes linéaires, qui sont appelées *espérances* (§ 2 ci-dessous), on emploie les mêmes termes qu'en description statistique : on parlera ainsi de la covariance de deux aléas numériques, de leur coefficient de corrélation linéaire, de la variance et de l'écart type d'un aléa numérique (cf. § 3 ci-dessous).

Dans les applications, il faut prendre bien garde aux erreurs d'interprétation que peut entraîner cette homonymie. En particulier, en inférence statistique, seront constamment présents les mêmes mots, par exemple le mot variance, pour désigner un nombre associé à la description des données expérimentales de l'échantillon : *variance de l'échan-*

tillon, et d'autre part pour désigner un nombre relatif *au modèle*, c'est-à-dire à un aléa numérique : variance de l'aléa numérique dont les observations sont supposées être un échantillon.

Le tableau ci-dessous récapitule les équivalences.



— Dans le même ordre d'idées, on ne calculera pas dans ce manuel l'expression générale de la distribution de probabilité de la somme de deux aléas numériques f et g finis et indépendants, de distributions de probabilité connues. Les outils de l'algèbre linéaire permettent en effet de se passer entièrement de cette expression dans les questions traitées ici.

*e) *Combinaisons linéaires d'aléas numériques*

(Ω, p) étant donné, si à chaque aléa numérique f on associe l'ensemble quotient Ω/f, et, pour chaque valeur $y \in \mathbf{R}$ prise par f, la classe correspondante $\Omega_y = \{ x : x \in \Omega, fx = y \}$, la fonction $f : \Omega \to \mathbf{R}$ peut s'écrire comme combinaison linéaire :

$$f = \sum_{y \in f\Omega} y \varphi_{\Omega_y}$$

où φ_{Ω_y} est la *fonction caractéristique* (*MSH*, chap. V, § 4) de la partie Ω_y de Ω ($\varphi_{\Omega_y}(x) = 1 \Leftrightarrow x \in \Omega_y$, et vaut 0 sinon).

A Ω_y est associé l'aléa de Bernoulli B_y :

B_y	0	1
	$(1 - p(\Omega_y))$	$p(\Omega_y)$

f, en tant qu'aléa numérique défini sur (Ω, p), peut donc être considéré comme combinaison (avec les coefficients y) des aléas de Bernoulli B_y.

D'ailleurs, si $\Omega = \{ 1, 2, 3, \ldots, n \}$ possède n éléments, une *base* du vectoriel $\mathbf{R}^\Omega$ est constituée par les n fonctions caractéristiques $\varphi_{\{i\}}$: les vecteurs qui s'écrivent, par la suite de leurs coordonnées, $(1, 0, \ldots, 0)$; $(0, 1, 0, \ldots, 0)$; etc.

Une *base* du vectoriel des aléas numériques définis sur (Ω, p) est constituée par les n aléas de Bernoulli B_i :

B_i	0	1
	$(1 - p_i)$	p_i

où $\mathbf{p} = (p_1, p_2, \ldots, p_n)$

Cette décomposition d'un aléa numérique fini en combinaison linéaire d'aléas de Bernoulli n'implique rien quant à l'indépendance en probabilité des aléas de Bernoulli composant la somme.

2. Espérances

a) *Définition de l'espérance*

A chaque application $f \in \mathbf{R}^{\Omega}$ de Ω dans $\mathbf{R}$, faisons correspondre la somme des valeurs fx prises, lorsque x parcourt Ω, par son image, pondérées par leurs probabilités p_x ; on définit ainsi un nombre, noté $\mathrm{E}_p(f)$, et appelé l'*espérance* de l'aléa numérique f. On peut écrire :

$$\boxed{\mathrm{E}_p(f) = \sum_{x \in \Omega} p_x f x}$$

Exemples :

1. $\Omega = \{a, e, i, o, u\}$ $\quad \mathbf{p} = (1/10, 1/2, 1/10, 1/5, 1/10)$.

$f =$

a	e	i	o	u
1	-1	2	1	2

$$\mathrm{E}_p(f) = (1/10) \times 1 - (1/2) \times 1 + (1/10) \times 2 + (1/5) \times 1 + (1/10) \times 2 = 1/5$$

2. *Espérance d'un aléa de Bernoulli.*

L'aléa de Bernoulli de paramètre θ a pour espérance :

$$0 \times (1 - \theta) + 1 \times \theta = \theta$$

L'espérance d'un aléa de Bernoulli est égale à la valeur du paramètre, c'est-à-dire la probabilité pour que l'aléa prenne la valeur 1.

L'*espérance* E_p qui à chaque application numérique $f : \Omega \to \mathbf{R}$, c'est-à-dire à chaque *vecteur* du vectoriel $\mathbf{R}^{\Omega}$

fait correspondre un nombre, est une application de $\mathbf{R}^{\Omega}$ dans $\mathbf{R}$:

$$\mathrm{E}_p : \mathbf{R}^{\Omega} \to \mathbf{R}$$

Le vecteur f a pour coordonnées la suite des images des éléments de Ω ; l'espérance de f, somme pondérée de ses coordonnées, définit donc (cf. *MSH*, chap. XIII) une *application linéaire* du vectoriel $\mathbf{R}^{\Omega}$ dans $\mathbf{R}$, c'est-à-dire une *forme linéaire* sur $\mathbf{R}^{\Omega}$.

Comme les coefficients de cette *forme linéaire* constituent une distribution de probabilité sur Ω, ils sont tous positifs ou nuls, et de somme unité : E_p est donc même une *moyenne pondérée*, ou *forme linéaire positive normée.*

Espérance d'un aléa *numérique* image d'un aléa (Ω, p) : moyenne des valeurs prises par cet aléa pondérées par leurs probabilités.

$$\mathrm{E}_p(f) = \sum_{x \in \Omega} p_x f x$$

Synonymes :

Espérance mathématique ; valeur moyenne ; moment d'ordre un (de l'aléa numérique f).

b) *Propriétés de l'espérance*

L'espérance est une forme linéaire positive et normée, *i.e.* une moyenne pondérée ; c'est dire qu'elle jouit des propriétés :

1. $\mathrm{E}_p(f+g) = \mathrm{E}_p(f) + \mathrm{E}_p(g)$.

L'espérance d'une somme d'aléas numériques est la somme de leurs espérances (que ces aléas soient indépendants ou non).

Exemple :

Soit deux aléas de Bernoulli, de paramètres θ et θ' respectivement, images d'un même aléa $(\Omega, \mathbf{p})$. Ils ont respectivement θ et θ' pour espérance.

Leur somme a donc pour espérance $\theta + \theta'$.

2. $\mathrm{E}_p(\lambda f) = \lambda \mathrm{E}_p(f)$.

Multiplier un aléa numérique par un nombre λ multiplie son espérance par ce nombre.

Exemple :

Le produit de l'aléa de Bernoulli :

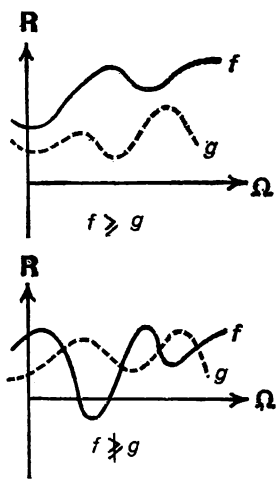
0	1
$1-\theta$	θ

par le nombre $3/2$ a pour espérance $3/2\,\theta$.

3. Si un aléa numérique a des valeurs toutes positives ou nulles : $\forall x \in \Omega$, $fx \geqslant 0$, son espérance est positive ou nulle. Il en résulte, grâce à la linéarité, que si un aléa numérique image par f de $(\Omega, \mathbf{p})$ est supérieur ou égal à l'image par g de $(\Omega, \mathbf{p})$, en ce sens que (on dit parfois dans ce cas que f est *uniformément* supérieur ou égal à g ; voir graphiques ci-contre) :

$$\forall x \in \Omega \qquad fx \geqslant gx$$

on a : $\mathrm{E}_p(f) \geqslant \mathrm{E}_p(g)$.



En effet, on peut écrire dans ce cas :

$$\forall x \in \Omega \qquad fx - gx \geqslant 0 \qquad \text{donc } E_p(f-g) \geqslant 0$$

Or :

$$E_p(f-g) = E_p(f) - E_p(g)$$

Donc :

$$E_p(f) - E_p(g) \geqslant 0$$

L'espérance est donc monotone :

$$f \geqslant g \; (\forall x \in \Omega, fx \geqslant gx) \Rightarrow E_p(f) \geqslant E_p(g)$$

4. L'espérance d'un aléa numérique *certain* est égale à la valeur certaine prise par cet aléa ; soit en effet f une application constante de Ω dans $\mathbf{R}$:

$$\forall x \in \Omega \qquad fx = a$$

$$E_p(f) = \sum_{x \in \Omega} p_x fx = \sum_{x \in \Omega} a p_x = a \sum_{x \in \Omega} p_x = a \times 1 = a$$

Exemple :

L'image d'un aléa numérique f, d'espérance $E_p(f)$ donnée, par une application affine :

$$t \rightsquigarrow \lambda t + a$$

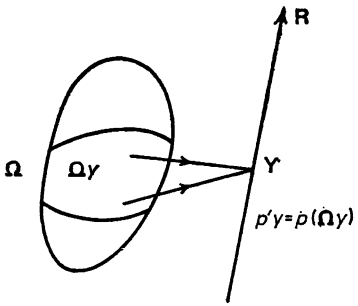
a pour espérance :

$$\lambda E_p(f) + a$$

c) *Calcul de l'espérance par les images*

L'aléa (Ω, p) a pour image par une application $f : \Omega \to \mathbf{R}$ un aléa numérique $(f\Omega, \mathbf{p}')$, où $p' = fp$ est la mesure image de p, et $f\Omega$, support de p', est l'image de Ω. Pour tout $y \in f\Omega$, $p'_y = p(\Omega y)$, où Ωy est la classe des éléments x de Ω ayant pour y image :

$$\Omega y = f^{-1} y = \{ x : x \in \Omega, fx = y \}$$



L'espérance $E_p(f)$ de l'aléa numérique $(f\Omega, p')$ image de (Ω, p) par f, somme lorsque x parcourt Ω des valeurs fx pondérées par les probabilités p_x, peut encore s'écrire, en regroupant dans la somme tous les éléments x de Ω appartenant à la même classe Ωy, et pour lesquels la valeur $fx\,(=y)$ est constante :

$$E_p(f) = \sum_{y \in f\Omega} y p'_y = \sum_{y \in f\Omega} y p(\Omega y)$$

Exemple :

On reprend le premier du § 2-*a supra*.

On a : $f\Omega = \{-1, 2, 1\}$;

$\Omega_1 = \{e\}$ $\Omega_1 = \{a, o\}$;

$\Omega_2 = \{i, u\}$.

$\mathbf{p}' = (p'_{-1}, p'_1, p'_2)$

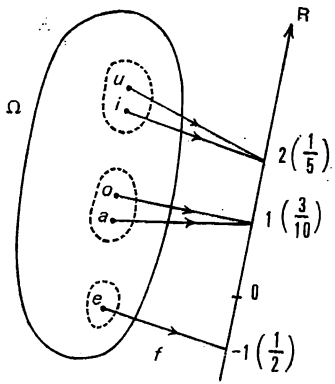
$p'_{-1} = p(\{e\}) = 1/2$

$p'_1 = p(\{a\}) + p(\{o\}) = 1/10 + 1/5 = 3/10$

$p'_2 = p(\{i\}) + p(\{u\}) = 1/10 + 1/10 = 1/5$

D'où :

$$E_p(f) = 1/2 \times (-1) + 3/10 \times 1 + 1/5 \times 2 = 1/5$$



*Ce second mode de calcul de l'espérance, au moyen des valeurs et des probabilités *image*, est d'ailleurs une conséquence de ce que tout aléa numérique fini est combinaison linéaire d'aléas de Bernoulli (§ 1-*e*), et de la linéarité de l'espérance ; dans les notations de 1-*e* :

$$f = \sum_{y \in f\Omega} y\varphi_{\Omega y} \Rightarrow \mathrm{E}_p(f) = \sum_{y \in f\Omega} y\mathrm{E}_p(\varphi_{\Omega y})$$

Or :

$$\mathrm{E}_p(\varphi_{\Omega y}) = 0 \times (1 - p(\Omega y)) + 1 \times p(\Omega y) = p(\Omega y)$$

On retiendra qu'il y a toujours deux façons *équivalentes* de calculer l'espérance d'un aléa numérique image d'un aléa (Ω, p) :
— en faisant la sommation par rapport au *départ* :

$$\boxed{\sum_{x \in \Omega} p_x f x}$$

On *somme pour tous les* x *de* Ω ;
— en faisant la sommation par rapport aux *images* :

$$\boxed{\sum_{y \in f\Omega} y p(\Omega y)}$$

On somme pour tous les y de $f\Omega$: valeurs images y pondérées par leurs probabilités.

d) *Les probabilités comme espérances*

Soit Ω un ensemble, p une mesure de probabilité sur Ω ; chaque partie $A \subset \Omega$ a pour probabilité :

$$p(A) = \sum_{x \in A} p_x$$

Associons à chaque partie A de Ω sa fonction caractéristique :

$$\varphi_A : \Omega \to \mathbf{R}$$

$$\varphi_A x = 1 \Leftrightarrow x \in A ; \qquad \varphi_A x = 0 \Leftrightarrow x \notin A$$

L'aléa image de (Ω, p) par φ_A est l'aléa de Bernoulli :

0	1
$(1 - p(A))$	$p(A)$

Il a donc pour espérance :

$$E_p(\varphi_A) = 0 \times (1 - p(A)) + 1 \times p(A) = p(A)$$

La probabilité d'une partie $A \subset \Omega$ est donc l'*espérance* de l'aléa de Bernoulli image de Ω par la fonction caractéristique de A.

*Les propriétés des mesures de probabilités peuvent donc s'exprimer comme des propriétés de l'espérance ; en particulier :

— Si $A = \Omega$, $\varphi_A = \varphi_\Omega$ est la fonction constante qui vaut 1 partout ; $E_p(\varphi_\Omega) = 1$.

D'où : $p(\Omega) = 1$.

— Si $A = \varnothing$, $\varphi_A = \varphi_\varnothing$ est la fonction constante qui vaut 0 partout ; $E_p(\varphi_\varnothing) = 0$.

D'où : $p(\varnothing) = 0$.

— Si A et B sont deux parties *disjointes*, on a (vérifiez-le) :

$$\varphi_{A \cup B} = \varphi_A + \varphi_B$$

Donc :

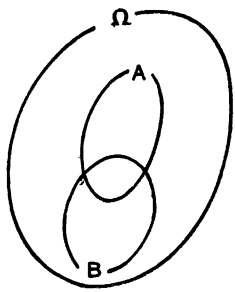
$$E_p(\varphi_{A \cup B}) = E_p(\varphi_A) + E_p(\varphi_B)$$

C'est-à-dire :

$$p(A \cup B) = p(A) + p(B)$$

— Plus généralement, si A et B sont des parties quelconques on a :

$$\varphi_{A \cup B} + \varphi_{A \cap B} = \varphi_A + \varphi_B$$



Les deux membres valent o dans $A' \cap B'$ (A' = complémentaire de A dans Ω), 1 dans leur différence symétrique, et 2 dans leur intersection ; l'espérance étant *linéaire*, on a :

$$E_p(\varphi_{A \cup B} + \varphi_{A \cap B}) = E_p(\varphi_{A \cup B}) + E_p(\varphi_{A \cap B})$$
$$E_p(\varphi_A + \varphi_B) = E_p(\varphi_A) + E_p(\varphi_B)$$

D'où :

$$\boxed{P(A \cup B) + p(A \cap B) = p(A) + p(B)}$$

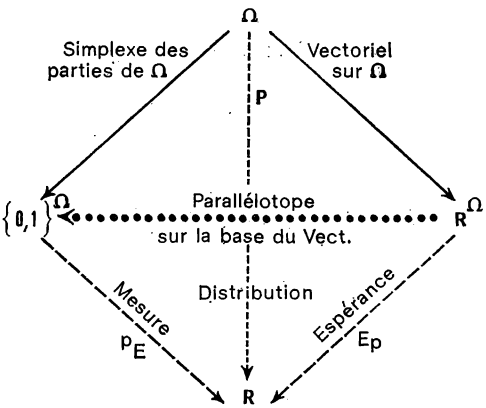
On voit ainsi que si les règles du calcul des probabilités, permettent de montrer, lorsqu'on les adopte, que l'espérance est une forme linéaire positive normée sur $\mathbf{R}^\Omega$, elles peuvent inversement être déduites des propriétés des formes linéaires. Plus précisément, à toute forme linéaire positive normée E sur le vectoriel $\mathbf{R}^\Omega$ des

applications numériques de Ω dans $\mathbf{R}$ correspond une mesure de probabilité p_E sur Ω définie par :

$$\forall A \subset \Omega \qquad p_E(A) = E(\varphi_A)$$

Le calcul des probabilités peut ainsi être inclus dans le cadre plus vaste de l'algèbre linéaire.

Un diagramme peut résumer les rapports entre les structures en jeu — Ω, le simplexe de ses parties, le corps $\mathbf{R}$ des réels, le vectoriel $\mathbf{R}^\Omega$ — et entre distribution, mesure et forme linéaire (espérance). Cf. aussi *MSH*, chap. XII, § 3-*b*.



Exercice :

Prouver que si $E : \mathbf{R}^\Omega \to \mathbf{R}$ a les propriétés :

Linéarité : $E(f+g) = E(f) + E(g)$,
$E(\lambda f) = \lambda E(f)$;

Positivité : $\forall x \in \Omega, fx \geqslant 0 \Rightarrow E(f) \geqslant 0$;

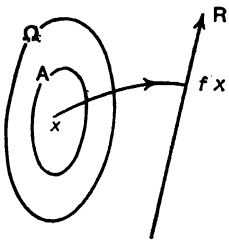
Être normée : $\forall x \in \Omega, fx = a \Rightarrow E(f) = a \; (a \in \mathbf{R})$,

alors p_E définie sur l'ensemble $P(\Omega)$ des parties de Ω par : $p_E(A) = E(\varphi_A)$ où φ_A est la fonction caractéristique de la partie A, est une mesure de probabilité.

e) *Espérances conditionnelles*

A chaque partie $A \subset \Omega$ on peut associer la mesure de probabilité conditionnelle (si A), p_A, définie comme au chapitre VIII, § 1 :

$$\forall B \subset \Omega,\ p_A(B) = \frac{p_\Omega(A \cap B)}{p_\Omega(A)} = \frac{p(A \cap B)}{p(A)} \qquad (p(A) > 0)$$



L'espérance conditionnelle (si A) d'un aléa numérique *f défini sur* Ω se calcule comme l'espérance sur Ω mais en prenant la restriction de f à A et en pondérant par la distribution de probabilité p_A conditionnelle (si A). Si l'on note cette espérance conditionnelle $E_{p_A}(f)$:

$$E_{p_A}(f) = \sum_{x \in A} p_A(x) fx \quad (p_A(x) = p_A(\{x\}))$$

Soit A' la partie complémentaire de A dans Ω ; si l'on suppose qu'elle a également une probabilité non nulle pour la mesure p :

$$p(A') = 1 - p(A) > 0$$

on peut définir de même l'espérance conditionnelle (si A') de f :

$$E_{p_{A'}}(f) = \sum_{x \in A'} p_{A'}(x) fx$$

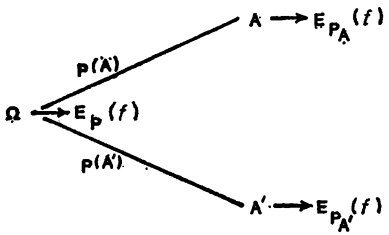
Comme :

$$\forall x \in A,\quad p_A(x) = \frac{p_x}{p(A)} \quad \text{et} \quad \forall x \in A',\quad p_{A'}(x) = \frac{p_x}{p(A')}$$

On voit par substitution que :

$$\mathrm{E}_p(f) = p(\mathrm{A})\,\mathrm{E}_{p_\mathrm{A}}(f) + p(\mathrm{A}')\,\mathrm{E}_{p_{\mathrm{A}'}}(f)$$

C'est-à-dire que l'espérance (sur Ω et par rapport à la mesure p) d'un aléa numérique f se calcule comme l'espérance d'un aléa prenant *deux valeurs*, les espérances condi-



tionnelles si A et si A′ respectivement, avec les probabilités $p(\mathrm{A})$ et $p(\mathrm{A}')$ respectivement.

Plus généralement, à toute partition de Ω en classes A_1, A_2, ..., A_k, sont associées k espérances conditionnelles $\mathrm{E}_{p_{\mathrm{A}_i}}$, et l'on a pour tout aléa numérique f sur Ω :

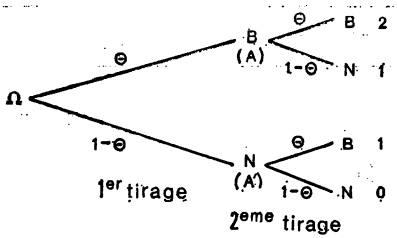
$$\mathrm{E}_p(f) = \sum_{i=1}^{k} p(\mathrm{A}_i)\,\mathrm{E}_{p_{\mathrm{A}_i}}(f)$$

Exemples :

1. Soit deux tirages *indépendants* (1) dans une urne dont la proportion des boules blanches est θ. Les résultats possibles

(1) *i.e.*, avec remise.

et le nombre total des blanches sont figurés sur l'arbre ci-dessous.



Si la première boule tirée est B (événement A), l'espérance *conditionnelle* (si A) du nombre de blanches est :

$$2\theta + 1 \times (1 - \theta) = 1 + \theta$$

Si la première boule tirée est N (événement A'), l'espérance conditionnelle est :

$$1 \times \theta + 0 \times (1 - \theta) = \theta$$

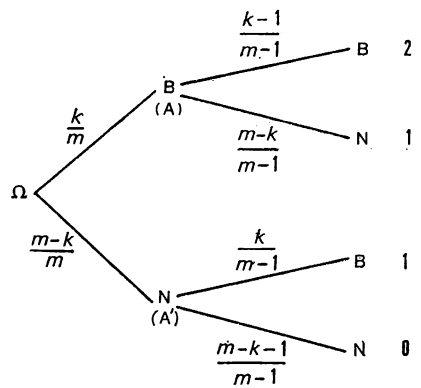
D'où pour l'espérance du nombre des blanches sur l'ensemble des deux tirages :

$$\theta (1 + \theta) + (1 - \theta)\theta = 2\theta$$

Si, au lieu de 2 boules, on en tire n, on montre de même par récurrence que l'espérance est $n\theta$. On établira plus loin ce résultat par un autre mode de calcul (§ 2-*f*).

2. *Espérance d'un aléa hypergéométrique.*

Si les deux tirages ne sont plus indépendants, mais sans remise, et si l'on suppose que l'urne contient m boules, dont k sont blanches (notations du § 4, chap. VII), le schéma devient :



Les deux espérances conditionnelles sont :

Si A (première boule tirée blanche) :

$$2\,\frac{k-1}{m-1}+1\,\frac{m-k}{m-1}=\frac{m+k-2}{m-1}$$

Si A′ (première boule tirée noire) :

$$1\,\frac{k}{m-1}+0\,\frac{m-k-1}{m-1}=\frac{k}{m-1}$$

D'où pour l'espérance du nombre de boules blanches dans la suite des deux tirages :

$$\frac{k}{m}\times\frac{m+k-2}{m-1}+\frac{m-k}{m}\times\frac{k}{m-1}$$

$$=\frac{k}{m(m-1)}(2m-2)=2\,\frac{k}{m}$$

L'espérance *est encore égale au double* de la proportion *initiale* $\frac{k}{m}$ des boules blanches dans l'urne.

Si au lieu de tirer deux boules, sans remise, on en tire n, le nombre de boules blanches tirées constitue, on l'a dit au chapitre VII, § 4, un aléa hypergéométrique.

On peut montrer, comme *supra*, par récurrence sur n, que l'espérance de cet aléa est :

$$n\frac{k}{m}$$

L'espérance est égale au nombre de boules tirées (la taille de l'échantillon) multiplié par la proportion *initiale* des boules blanches dans l'urne.

f) *Application de la linéarité de l'espérance à l'aléa binomial*

L'aléa binomial de paramètre n et θ (chap. VII, § 6 et chap. VIII, § 3-*d*) est interprétable comme la fréquence des boules blanches dans une série de n tirages *indépendants* (avec remise) dans une urne ayant la proportion θ de boules blanches.

Cet aléa f numérique peut se décomposer en une somme :

$$f = f^1 + f^2 + \ldots + f^n$$

où f^i est l'aléa de Bernoulli qui vaut 1 si la i^e boule tirée est blanche, et 0 sinon. L'hypothèse faite, qu'à chaque tirage la composition de l'urne est la même (θ), signifie que chacun de ces aléas de Bernoulli est :

f^i	0	1
	$1-\theta$	θ

Chacun a donc θ pour espérance. L'espérance d'une somme étant la somme des espérances, il en résulte que l'espérance de l'aléa binomial est :

$$\underbrace{\theta + \theta + \theta + \ldots}_{n \text{ fois}} + \theta = n\theta$$

> L'*aléa binomial* de paramètres n et θ a pour espérance :
>
> $$n\theta$$

Exemple :

On procède à 100 tirages indépendants dans une population dont le quart présente le caractère α ; l'espérance est :

$$100 \times 1/4 = 25$$

Espérance de la fréquence relative dans une suite de tirages indépendants.

La *fréquence relative* est le rapport de la fréquence, dans une série de n tirages, au nombre n d'individus tirés : si l'on procède à 20 tirages, et que 12 individus tirés présentent le caractère α, la fréquence est 12, et la fréquence relative est :

$$12/20 = 3/5 = 60/100$$

La fréquence relative est donc un aléa numérique déduit de l'aléa binomial en divisant celui-ci par n. Or on a pour l'aléa binomial :

$$E_p(f) = n\theta$$

Mais, en vertu de la linéarité de l'espérance $(\mathrm{E}_p(\lambda f)) = \lambda \mathrm{E}_p(f))$:

$$\mathrm{E}_p(f/n) = 1/n\, \mathrm{E}_p(f) = n\theta/n = \theta$$

On remarque que cette espérance est la même *quel que soit le nombre d'éléments* tirés, et est égale à la *composition de l'urne.*

3. Covariance, variance

a) *Produit de deux aléas numériques*

Si f et g sont deux applications d'un même ensemble Ω dans $\mathbf{R}$, le produit fg de ces deux applications est par définition l'application $h : \Omega \to \mathbf{R}$ telle que :

$$\forall x \in \Omega \qquad hx = fx \times gx$$

L'image de chaque élément x de Ω est le produit des images de x par f et g respectivement.

Le produit de deux applications numériques est défini comme leur somme, en faisant jouer à la multiplication le rôle que joue l'addition dans la somme.

Exemple :

f est l'application qui fait correspondre 1 à tous les éléments d'une partie A de Ω, et 0 aux autres (f est la fonction caractéristique de A) ; g est de même la fonction caractéristique d'une autre partie B de Ω.

Le couple des applications f et g induit sur Ω la partition obtenue par croisement des partitions $\{\mathrm{A}, \mathrm{A}'\}$, $\{\mathrm{B}, \mathrm{B}'\}$ induites par f et g respectivement, c'est-à-dire la partition :

$$\{\mathrm{A} \cap \mathrm{B}, \mathrm{A} \cap \mathrm{B}', \mathrm{A}' \cap \mathrm{B}, \mathrm{A}' \cap \mathrm{B}'\}$$

Le produit fg vaut 1 pour tous les éléments de $A \cap B$, et 0 partout ailleurs.

Ω/g \ Ω/f	A : 1	A' : 0
B : 1	$A \cap B$: 1	$A' \cap B$: 0
B' : 0	$A \cap B'$: 0	$A' \cap B'$: 0

Si Ω est muni d'une distribution de probabilité $\mathbf{p}$, les images par f et g de l'aléa $(\Omega, \mathbf{p})$ sont deux aléas numériques, le produit de ces deux aléas numériques est *par définition* l'aléa numérique image de $(\Omega, \mathbf{p})$ par l'application fg produit de f et g.

Dans l'exemple, on a les deux aléas de Bernoulli :

f

0	1
$1 - p(A)$	$p(A)$

g

0	1
$1 - p(B)$	$p(B)$

Leur produit est l'aléa de Bernoulli :

fg

0	1
$1 - p(A \cap B)$	$p(A \cap B)$

Attention :

$p(A \cap B)$ n'est en général pas égal au produit $p(A)\,p(B)$ (cf. chap. VIII, § 3).

Produit de deux aléas numériques images par les applications f et g d'un *même* aléa (Ω, p) : aléa image de $(\Omega, \mathbf{p})$ par le *produit* fg des applications f et g.

Remarques :

On voit que, *mutatis mutandis*, le produit de deux aléas est défini comme la somme.

D'autre part on ne confondra pas ce produit avec le *produit cartésien* d'aléas défini chap. VIII, § 2.

b) *Espérance du produit de deux aléas numériques : un produit scalaire. Moments d'un aléa numérique*

On a vu au § 2-*b* que l'espérance constitue, sur le vectoriel des aléas numériques images de (Ω, p) dans $\mathbf{R}$, une forme linéaire, positive, normée. Il en résulte que l'espérance du produit fg de deux aléas numériques f et g, définie par :

$$\mathrm{E}_p(fg) = \sum_{x \in \Omega} p_x f x \, g x$$

est un *produit scalaire* sur ce vectoriel (chap. III, § 1 et 2) ; c'est une forme bilinéaire, symétrique, positive, normée (chap. III, § 2), définie comme aux § 2-*b*, 2-*c* du chapitre III (1).

L'espérance du produit jouit donc des propriétés du pro-

(1) On pourrait donc noter comme au chapitre III, $f.g$ l'espérance du produit de deux aléas numériques f et g. On a préféré ici la notation plus traditionnelle $\mathrm{E}_p(fg)$, qui a l'avantage de rappeler que ce produit scalaire est défini par la donnée de la mesure de probabilité p.

D'autre part la positivité n'est ici *stricte* comme pour le produit scalaire défini au chap. III que si p a Ω pour support ($\forall x \in \Omega, p_x > 0$), ce qu'on peut toujours supposer avec un choix convenable de Ω *fini*.

duit scalaire, et obéit aux mêmes règles de calcul ; en particulier :

$$\mathrm{E}_p\,[f\,(g+h)] = \mathrm{E}_p\,(fg) + \mathrm{E}_p\,(fh)$$
$$\mathrm{E}_p\,(\lambda f\,\mu g) = \lambda\mu\ \mathrm{E}_p\,(fg)$$

Ceci sera exploité au § 4 ci-dessous.

Lorsque $f = g$, l'espérance du produit devient :

$$\boxed{\mathrm{E}_p\,(f^2) = \sum_{x\in\Omega} p_x f^2 x}$$

C'est l'espérance du carré de l'aléa numérique f, souvent appelée *moment d'ordre* 2, et noté $m_2\,(f)$, de cet aléa.

$$\boxed{m_2\,(f) = \mathrm{E}_p\,(f^2)}$$

On peut définir de même les *moments d'ordre* 3, 4, etc.

$$\boxed{m_3\,(f) = \mathrm{E}_p\,(f^3) = \sum_{x\in\Omega} p_x f^3 x}$$

$m_3\,(f)$ est l'espérance du cube de l'aléa numérique f.

Dans la logique de ces notations, on peut considérer l'espérance de f comme *son moment d'ordre 1*, et la noter $m_1\,(f)$; cette notation est parfois employée.

Quant au moment d'ordre 2, il joue, comme on l'a vu (chap. IV, § 5), le rôle du *carré d'une distance* (distance euclidienne) sur le vectoriel des aléas numériques. En particulier, il permettra de définir la dispersion d'un aléa numérique au moyen de sa variance (§ 4-*b*).

c) *Aléa certain associé à l'espérance d'un aléa numérique*

Soit $E_p(f)$ l'espérance de l'aléa numérique image de l'aléa (Ω, p) par l'application $f : \Omega \to \mathbf{R}$. On désigne dans la suite par $\hat{f}$ l'application *constante* de Ω dans $\mathbf{R}$ qui envoie tous les éléments de Ω sur le nombre $E_p(f)$: l'image de (Ω, p) par $\hat{f}$ est l'aléa certain qui vaut $E_p(f)$ avec la probabilité 1 (cf. § 1-*b*).

Puisque $\hat{f}$ est une application constante, l'espérance de $\hat{f}$ est égale à la valeur constante prise :

$$E_p(\hat{f}) = E_p(f)$$

Exemple :

Pour un aléa binomial de paramètres n et θ, l'aléa certain qui lui est ainsi associé vaut $n\theta$.

Remarque :

L'aléa certain $\hat{f}$ associé à un aléa numérique f est, pour cet aléa, considéré comme vecteur, son *vecteur moyen* (celui dont toutes les coordonnées sont égales à la moyenne, ici l'espérance). Il joue par rapport aux aléas numériques le rôle que joue le vecteur moyen en description statistique, en particulier pour la définition de la covariance et de la variance (chap. III, § 5).

d) *Covariance de deux aléas numériques*

La *covariance* de deux aléas numériques images par deux applications f et g d'un même aléa (Ω, p) est par définition *l'espérance du produit* $((f - \hat{f})(g - \hat{g}))$; on la note $\mathrm{Cov}(f, g)$:

$$\mathrm{Cov}(f, g) = E_p((f - \hat{f})(g - \hat{g}))$$

$\hat{f}$ et $\hat{g}$ sont les aléas certains définis ci-dessus.

La covariance de deux aléas numériques est donc définie, au moyen du produit scalaire $E_p(fg)$, comme la covariance de deux vecteurs en description statistique (chap. III, § 5-*a*). Le même calcul formel qu'au § 5-*b*, chapitre III, permet de déduire de l'égalité :

$$(f-\hat{f})(g-\hat{g}) = fg - \hat{f}g - \hat{g}f + \hat{f}\hat{g}$$

l'égalité :

$$\boxed{\mathrm{Cov}(f,g) = E_p(fg) - E_p(f)\,E_p(g)}$$

La covariance de deux aléas numériques est égale à l'*espérance de leur produit moins le produit de leurs espérances.*

Exemple :

Dans l'exemple, qui a servi au § 4-*a*, de deux aléas de Bernoulli, on a :

$$E_p(f) = p(A) \qquad E_p(g) = p(B)$$

et puisque le produit est un aléa de Bernoulli de distribution de probabilité $(1 - p(A \cap B), p(A \cap B))$:

$$E_p(fg) = p(A \cap B)$$

Donc la covariance est :

$$\boxed{p(A \cap B) - p(A)\,p(B)}$$

On voit qu'en particulier elle est nulle si $p(A \cap B) = p(A)\,p(B)$, c'est-à-dire si les aléas f et g sont indépendants.

On admettra que ce résultat est général :

> **La covariance de deux aléas indépendants est nulle.**

Ce qui se formule encore :

> Si deux aléas sont indépendants, l'espérance de leur produit est égale au produit de leurs espérances.

Attention, la réciproque n'est pas vraie : la covariance de deux aléas numériques peut être nulle sans que ces aléas soient indépendants (p. ex. si $f = g^2$; cf. chap. V, § 2-*d*).

e) *Variance et écart type*

> La *variance* d'un aléa numérique f est par définition la covariance de cet aléa avec lui-même. On note :
>
> $$\mathrm{Var}\,(f) = \mathrm{E}_p\,((f - \hat{f})^2)$$

On obtient donc la variance de l'aléa numérique image de $(\Omega, \mathbf{p})$ par f en posant $f = g$ dans l'expression de la covariance ; en particulier, l'expression : $\mathrm{Cov}\,(f, g) = \mathrm{E}_p\,(fg) - \mathrm{E}_p\,(f)\,\mathrm{E}_p\,(g)$ de la covariance fournit pour la variance :

$$\mathrm{Var}\,(f) = \mathrm{E}_p\,(f^2) - (\mathrm{E}_p\,(f))^2$$

> La variance d'un aléa numérique est égale à l'*espérance de son carré* moins le *carré de son espérance.*

Exemple :

Pour un aléa de Bernoulli de paramètre θ :

0	1
$1 - \theta$	θ

le carré vaut 0 si et seulement si l'aléa vaut 0 et vaut 1 sinon ; donc un aléa de Bernoulli est son propre carré. L'espérance de son carré est donc son espérance θ, et sa variance est :

$$\theta - \theta^2 = \theta (1 - \theta)$$

La *variance d'un aléa de Bernoulli* est égale au produit des probabilités des deux valeurs 0 et 1 possibles de l'aléa :

$$\theta (1 - \theta)$$

L'*écart type* d'un aléa numérique f, noté $\sigma(f)$, est la *racine carrée de sa variance :*

$$\sigma(f) = [V(f)]^{1/2}$$

De même qu'en description statistique, la variance et l'écart type fournissent des mesures de la *dispersion d'un aléa* numérique ; on verra au § 5 (cf. également exercice ci-dessous) en quel sens précis ; l'écart type est homogène à une longueur, la variance au carré d'une longueur.

Exercice :

Prouvez que la variance d'un aléa fini est nulle si et seulement si cet aléa est certain.

f) *Coefficient de corrélation linéaire de deux aléas numériques*

Le coefficient de corrélation linéaire de deux aléas numériques f et g se définit comme celui de deux vecteurs en description statistique, et se note $r(f,g)$: c'est le rapport de la covariance de f et g au produit de leurs écarts types (cf. chap. V, § 3) supposés non nuls (f et g ne sont pas des aléas certains).

$$r(f,g) = \frac{\text{Cov}(f,g)}{\sigma(f)\,\sigma(g)}$$

Si f et g sont indépendants, leur coefficient de corrélation linéaire est nul ; dans tous les cas, il est compris entre 1 et -1, et n'est égal à l'une de ces valeurs extrêmes que si f est fonction affine de g ($\forall x \in \Omega, fx = agx + b$).

g) *Variance d'une somme d'aléas numériques*

Si f et g sont deux aléas numériques définis à partir d'un même aléa (Ω, p), leur somme $f+g$ a pour variance (même calcul que chap. III, § 5-*d*) :

$$\text{Var}(f+g) = \text{Var}\,f + 2\,\text{Cov}(f,g) + \text{Var}\,g$$

Comme en description statistique, la variance d'une somme *n'est pas* la somme des variances : il faut ajouter le double de la covariance.

Mais si les deux aléas numériques sont *indépendants*, leur covariance est nulle, et la variance la somme des variances.

Une somme d'aléas numériques *indépendants* a pour variance la somme des variances de ces aléas.

Application : variance de l'aléa binomial.

On a vu (§ 2-*f*) que l'aléa binomial de paramètres n et θ est la somme de n aléas de Bernoulli *indépendants* de paramètres θ. Sa variance est donc n fois la variance $\theta(1-\theta)$ d'un aléa de Bernoulli.

L'aléa binomial a pour variance :

$$n\theta(1-\theta)$$

Si l'on s'intéresse, dans une somme de n aléas de Bernoulli de paramètre θ (aléa binomial), non plus à la fréquence mais à la fréquence *relative*, c'est-à-dire au produit par $1/n$ de l'aléa binomial, la variance est :

$$\frac{1}{n^2} n\theta(1-\theta) = \frac{\theta(1-\theta)}{n}$$

En effet, si l'on multiplie un aléa par un nombre λ, sa variance est multipliée par λ^2 (cf. la *remarque*, chap. III, § 5-*d*).

$$\text{Var}(\lambda f) = \lambda^2 \text{Var}(f)$$

En particulier, si $\lambda = 1/n$:

$$\text{Var}(1/nf) = 1/n^2 \text{Var}(f)$$

D'où le résultat donné ci-dessus, et qui peut encore s'énoncer :

> Pour une série de n tirages *indépendants* dans une population dont une proportion θ d'individus présente un caractère donné α, la fréquence relative des individus tirés présentant ce caractère a pour variance :
>
> $$\frac{\theta(1-\theta)}{n}$$

Il résulte de cette expression que plus le nombre n des tirages est grand, plus la variance de la fréquence relative est faible : elle tend vers zéro lorsque n augmente indéfiniment.

On a de même pour l'*écart type* de l'aléa binomial :

$$\sqrt{n\theta\,(1-\theta)}$$

Celui de la fréquence relative est :

$$\sqrt{\frac{\theta\,(1-\theta)}{n}}$$

Remarque :

Si les tirages ne sont plus indépendants mais avec remise, la variance de la somme n'est plus la somme des variances ; l'aléa obtenu est dans ce cas l'*aléa hypergéométrique* (chap. VII, § 4) dont la variance ne peut donc se calculer par le procédé ci-dessus : il faut ajouter à la somme des n variances le double de la somme des $n\,(n-1)$ covariances.

On donne ici, sans démonstration, le résultat. Si l'urne contient m boules, dont k blanches, et si l'échantillon est de taille n (on tire n boules sans remise), la variance du nombre de boules blanches est :

$$n\frac{k}{m}\left(1-\frac{k}{m}\right)\left(1-\frac{n-1}{m-1}\right)$$

On remarque que pour une même composition *initiale* ($\theta = k/m$) de l'urne, la variance de l'aléa hypergéométrique est plus faible que celle de l'aléa binomial pour une même taille de l'échantillon, puisqu'elle s'écrit alors :

$$(n\theta\,(1-\theta))\left(1-\frac{n-1}{m-1}\right)$$

C'est la variance de l'aléa binomial $n\theta\,(1-\theta)$ multipliée par le nombre (inférieur à 1) $1-\dfrac{n-1}{m-1}$.

On verra au § 4 comment les tirages dans une urne permettent d'estimer la proportion (inconnue) de boules blanches dans celle-ci, et comment la variance permet, en un certain sens, de mesurer la valeur de cette estimation ; l'écart entre variance de l'aléa binomial et variance de l'aléa hypergéométrique mesure la « perte d'information » qui résulte de la remise dans l'urne des boules tirées.

4. Première loi des grands nombres Inférence statistique

a) *Une inégalité remarquable*

— Soit $(\Omega, \mathbf{p})$ un aléa quelconque ; $f : \Omega \to \mathbf{R}$ définit un aléa numérique. Considérons d'autre part l'aléa de Bernoulli défini par $\varphi : \Omega \to \mathbf{R}$

$$\text{où } \varphi x = 0 \Leftrightarrow -\varepsilon \leqslant fx \leqslant \varepsilon$$
$$\varphi x = \varepsilon^2 \text{ sinon}$$

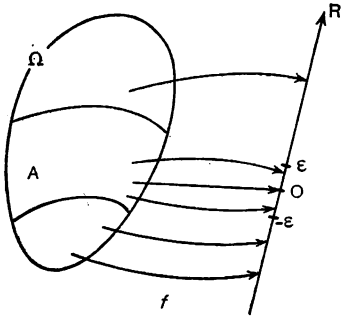
où ε est nombre donné.

Autrement dit, si l'on désigne par A la partie de Ω telle que :

$$x \in \mathrm{A} \Leftrightarrow -\varepsilon \leqslant fx \leqslant \varepsilon$$

φ vaut 0 sur A, et ε^2 sur sa partie complémentaire A'. C'est le produit par ε^2 de la fonction caractéristique de la partie A de Ω. L'aléa image de $(\Omega, \mathbf{p})$ par φ est :

0	ε^2
p (A)	p (A')



— D'après la définition même de φ, il est clair que :

$$\forall x \in \Omega \qquad (fx)^2 \geqslant \varphi x$$

si $x \in A$:

$$(fx)^2 \geqslant 0 = \varphi x$$

et si $x \in A'$:

$$(fx)^2 \geqslant \varepsilon^2 = \varphi x$$

On a donc d'après la *positivité* de l'espérance (§ 2-*b*), d'où résulte sa *monotonie* :

$$E_p(f^2) \geqslant E_p(\varphi)$$

Or :

$$E_p(\varphi) = \varepsilon^2 p(A')$$

Comme A' est la partie de Ω pour laquelle la valeur absolue de fx, soit $|fx|$, est supérieure à ε, on peut écrire :

$$p(A') = \Pr(|fx| > \varepsilon)$$

D'où l'inégalité :

$$\boxed{E_p(f^2) \geqslant \varepsilon^2 \Pr(|fx| > \varepsilon)}$$

b) *Variance et dispersion*
Inégalité de Bienaymé et Tchebicheff

L'inégalité que l'on vient d'établir est valable pour *tout* aléa numérique, et pour *tout* nombre ε ; en particulier, elle est valable pour l'aléa numérique $(f - \hat{f})$, où $\hat{f}$ est l'appli-

cation constante ayant $E_p(f)$ pour image. Comme chaque élément x de Ω a pour image dans cette application $(f - \hat{f})$ le nombre :

$$fx - \hat{fx} = fx - E_p(f)$$

Et comme, par définition de la variance :

$$E_p((f - \hat{f})^2) = \text{Var}(f)$$

on obtient :

$$\boxed{\text{Var}(f) \geqslant \varepsilon^2 \, \text{Pr}(|fx - E_p(f)| > \varepsilon)}$$

C'est l'inégalité dite de *Bienaymé et Tchebicheff*. Elle s'écrit encore :

$$\boxed{1/\varepsilon^2 \, \text{Var}(f) \geqslant \text{Pr}(|fx - E_p(f)| > \varepsilon)}$$

Sous cette forme, on va comprendre sa signification : le second membre, c'est la probabilité pour que l'aléa numérique f s'écarte de son espérance de plus de ε ; l'inégalité exprime que cette probabilité est bornée par $1/\varepsilon^2 \, \text{Var}(f)$, c'est-à-dire par un nombre *proportionnel à la variance*. On voit par là en quoi la variance est une mesure des fluctuations (aléatoires) d'un aléa numérique autour de son espérance, c'est-à-dire la dispersion de cet aléa.

Exemple :

Si un aléa a une variance égale à 5, la probabilité pour qu'il diffère de plus de 10 unités de son espérance est inférieure à :

$$5/10^2 = 5\ \%$$

Si sa variance est 1, cette probabilité est inférieure à :

$$1/10^2 = 1\ \%$$

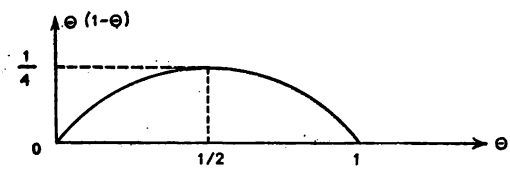
c) *Première loi des grands nombres*

Appliquons l'inégalité de Bienaymé et Tchebicheff à la fréquence relative dans une série de n tirages indépendants dans une population dont une proportion θ d'individus possède le caractère observé. On a vu que cette fréquence relative est un aléa dont l'espérance est θ et la variance $\frac{\theta(1-\theta)}{n}$. Désignons par ν/n la fréquence relative ; on a :

$$1/\varepsilon^2 \frac{\theta(1-\theta)}{n} \geqslant \Pr(|\nu/n - \theta| > \varepsilon)$$

Donc, pour une valeur *fixée* de ε, la probabilité pour que la fréquence relative diffère de la proportion θ de plus de ε tend vers zéro lorsque le nombre n des tirages augmente indéfiniment, et ceci quel que soit θ.

Le produit $\theta(1-\theta)$ est, lorsque θ varie de 0 à 1, compris entre 0 et 1/4 (1).



On peut donc encore écrire :

$$\Pr(|\nu/n - \theta| > \varepsilon) \leqslant 1/n \times 1/4\,\varepsilon^2$$

Exemple :

Prenons $\varepsilon = 1/10 = 0{,}1$; $1/\varepsilon^2 = 100$.
On a : $\Pr(|\nu/n - \theta| > 1/10) \leqslant 25/n$.

(1) La courbe est un arc de parabole.

Quel est le nombre de tirages à faire pour que la fréquence relative ait une probabilité inférieure à 5 % de différer de plus de 1/10 de la proportion θ ? Il faut que :

$$25/n = 5/100$$

C'est-à-dire $n = 500$.

Si l'on fait 500 tirages (avec remise), on est sûr qu'il y a au moins 5 chances sur 100 pour que la fréquence relative diffère de plus de 1/10 de la proportion θ. Si l'on fait 5 000 tirages, cette probabilité tombe à 5 sur 1 000, etc.

Le résultat obtenu peut s'énoncer en disant que la fréquence relative *converge en probabilité* vers la proportion θ des individus présentant le caractère étudié dans la population : *plus le nombre de tirages indépendants est grand, plus la probabilité, pour que l'écart entre fréquence relative observée sur l'échantillon et proportion dans la population soit supérieur à une valeur donnée, est petite. Cette probabilité tend vers zéro lorsque la taille* n *de l'échantillon augmente indéfiniment.*

La convergence en probabilité mise ici en évidence pour la fréquence relative d'un aléa binomial constitue ce qu'on appelle la *première loi des grands nombres*, ou *loi de Bernoulli.*

Remarque :

La quantité $\frac{\theta(1-\theta)}{n\varepsilon^2}$, et *a fortiori* la quantité $\frac{1}{4n\varepsilon^2}$ qui sert de borne supérieure à la valeur de la probabilité $\Pr(|\nu/n - \theta| > \varepsilon)$ est en général bien supérieure à cette probabilité. Cette borne est commode pour montrer la convergence, elle est très mauvaise en général pour estimer la probabilité en question, qui peut être beaucoup plus petite.

D'autre part (cf. remarque finale, § 4-*g*) on obtient une convergence plus rapide en faisant des tirages *sans remise*, et non avec remise comme il est supposé ici.

d) *Inférence statistique*

La première loi des grands nombres montre comment, grâce au *modèle probabiliste*, la description statistique permet de faire de l'*inférence* statistique. La fréquence relative des individus présentant un caractère donné dans un échantillon prélevé d'une population par tirages avec remise constitue en effet, lorsque le tirage est effectué et l'observation faite, une *description statistique* (*i.e.* par dénombrement) de l'observation.

Le *modèle probabiliste* consiste à considérer *avant le tirage* cette fréquence comme un aléa numérique : à construire d'une part l'ensemble Ω des $(n + 1)$ résultats possibles de l'expérience, et d'autre part à munir cet ensemble d'une mesure de probabilité (ici, la distribution binomiale). Le paramètre θ de cette distribution binomiale est, on l'a vu, si des précautions techniques dans l'échantillonnage sont prises, égal à la proportion, en général *inconnue* de l'expérimentateur, des individus présentant le caractère étudié *dans la population.*

Le théorème de convergence en probabilité montre que cette proportion peut être estimée, avec une certaine probabilité d'erreur, par la fréquence observée : procéder à cette estimation, c'est faire une *inférence* (statistique) ; à partir de l'observation (partielle) sur l'échantillon on porte un jugement probabiliste sur la population d'où il est extrait.

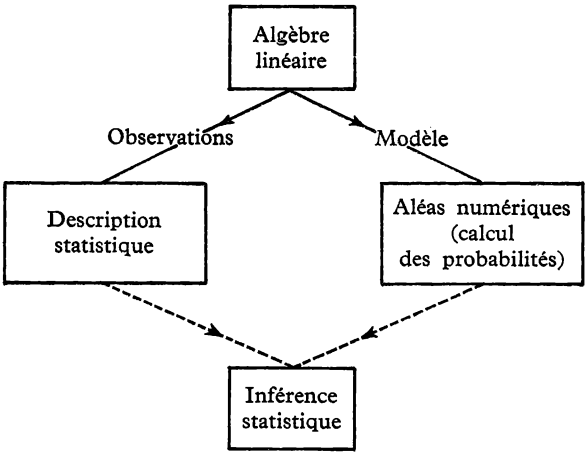
Remarque :

Bien entendu, à propos de cette estimation, on se gardera de parler de « fourchette », c'est-à-dire d'un intervalle dans lequel serait située la valeur inconnue de la proportion à estimer. Pourquoi cela est-il absurde ?

La démonstration du théorème de convergence en probabilité montre toute la puissance de l'outil que constitue

l'algèbre linéaire en calcul des probabilités ; dans tout ce qui précède, on n'a jamais eu à se servir effectivement de l'expression de la distribution binomiale, expression relativement compliquée. Tout a reposé sur la structure vectorielle des aléas numériques, et sur le fait que l'espérance est une forme linéaire positive et la covariance un produit scalaire.

L'algèbre linéaire constitue le cadre abstrait dont relève l'essentiel du calcul des probabilités, aussi bien que de la description statistique ; comme l'inférence statistique résulte de l'application de la première de ces interprétations de l'algèbre linéaire (calcul des probabilités) à la seconde (description statistique), ceci peut se résumer par le diagramme (cf. également § 1-*d*) :



INDEX DES SYMBOLES ET NOTATIONS

	Pages
$\simeq$	29
Σ	184
$\perp$	94
$\mathbf{N}$	14
$\mathbf{Z}$	16
$\mathbf{Q}$	19
$\mathbf{Q}^*$	19
$\mathbf{D}$	23
$\mathbf{R}$	22
$\mathbf{R}^*$	22
$\mathbf{R}_0^*$	113
ω	26
η	27
ζ	26
a^x	48
a^0	50
a^{-n}	51
$a^{1/n}$	52
$a^{p/q}$	52
$\log_a x$	57
$\mathbf{x}.\mathbf{y}$	83
$\|\mathbf{x}\|$	91
$\mathrm{V}_{\mathbf{x}}$	100
$m_{\mathbf{x}}$	97
$\mathbf{m}_{\mathbf{x}}$	98
$\sigma\mathbf{x}$	130
$\mathrm{Cov}_{\mathbf{xy}}$	98
$\mathrm{Cos}\,(\mathbf{x}, \mathbf{y})$	138
$r_{\mathbf{xy}}$	152
$\mathrm{E}_p(f)$	249
$\mathrm{V}(f)$	270
$\sigma(f)$	271
$\mathrm{Cov}(f, g)$	268
$r(f, g)$	272

INDEX DES TERMES

Abélien (groupe —), 17.
Additivité (des mesures), 171.
Affine (fonction —), 44.
Ajustement (droite d' —),156.
Aléa, 192.
Aléatoire (variable —), 192, 199.
Anneau, 19.
Approximation, 29.
Arithmétique (échelle —), 69.
Arithmétique (progression —), 45.
Associatif, 15.
Automorphisme, 43.

Base (d'une exponentielle), 48.
Base (d'un logarithme), 57.
BAYES (formule de —), 229.
BERNOULLI (aléa de —), 193.
BERNOULLI (Loide —), 279.
BIENAYMÉ (inégalité de — et TCHEBICHEV), 277.
Bilinéaire (forme —), 88.
Bi-logarithmique (échelles —s), 71.
Binôme (formule du —), 237.
Binomial (aléa —, distribution —e), 212.
BRAVAIS (coefficient de —, corrélation de — et PEARSON), 152.

Certain (aléa —), 241.
Certain (événement —), 167.
Commutatif, 15 ; (groupe —), 17.
Compatible, 18.
Conditionnel (espérance —le), 258; (probabilité — e), 220.
Continu (ordre —), 28.
Convergence (en probabilité), 279.
Corrélation (coefficient de —), 152, 272.
Corps, 22.
Cosinus, 138.
Covariance, 98, 268.
Croisement (d'aléas), 224.
Croissante (fonction —), 41.

Décimal (nombre —), 13.
Décroissante (fonction —), 42.
Dénombrable (ensemble —), 27.
Discret (ordre —), 27.
Distance, 132.
Distribution (de probabilité), 186.

Écart type, 130, 271.
Échelle, 29.
Élémentaire (événement —), 167.
Endomorphisme, 43.
Entier (nombre — naturel), 14 ; (nombre — relatif), 12.
Équiprobable, 189.
Erreur, 29.
Espérance, 250.
Euclidienne (distance —), 112 ; (longueur —), 91, 112.
Événement, 165.
Exponentiel (fonction —le), 46, 55.
Exposant, 62.

Forme (bilinéaire), 88 ; (— linéaire), 88.
Fréquence, 263.

Géométrique (progression —), 47.

Homothétie, 81.

Hypergéométrique (aléa —, distribution —), 212.

Image (aléa —), 199 ; (mesure —), 199.
Impossible (événement —), 167.
Incompatible (événements —s), 167.
Indépendant (aléas —s), 231 ; (événements —s), 233.
Inégalité (de BIENAYMÉ et TCHEBICHEFF), 277 ; (— de SCHWARZ), 106 ; (— du triangle), 107, 119.
Interquantile (intervalle —), 135.
Interquartile (intervalle —), 135.
Intervalle (échelles d'—s), 45.
Isomorphisme, 56.

KOLMOGOROFF (axiomes de —), 172.

Linéaire (coefficient de corrélation —), 152.
Linéaire (fonction —), 39 ; (forme —), 88.
Logarithme (fonction —), 57.
Longueur (euclidienne), 91, 112.

Marginal (distribution —e), 227.
Mesure, 177.
Métrique (espace —), 132.
Moment (—s d'un aléa), 267.
Monoïde, 16.
Monotone (fonction —), 41.
Moyen (vecteur —), 98.
Moyenne (d'un aléa), 250 ; (— d'un vecteur), 97.

Naturel (entier —), 14.
Norme, 112.
Numérique (aléa —), 241.

Ordinale (échelle —), 29.
Ordonné (anneau —), 19 ; (corps —), 22 ; (Groupe —), 18.
Ordonnée (à l'origine), 44.
Ordre (type d'—), 26.
Orthogonal, 94.

Pile ou face (aléa de —), 206 ; (— du produit scalaire), 87.
Probabilité (mesure de —), 172.
Produit (d'aléas numériques), 266.
Produit (cartésien — d'aléas), 224 ; (— scalaire), 83, 90.
Puissance (fonction —), 62.
PYTHAGORE (relation de —), 96.

Qualitative (échelle —), 29.
Quantitative (échelle —), 29.
Quotient (aléa —), 210.

Rationnel (nombre —), 12.
Réalisation (d'un événement), 166.
Réel (nombre —), 22.
Régression (droites de —), 157.
Relatif (entier —), 12.
Relative (fréquence —), 263.

Scalaire (produit), 83, 90.
SCHWARZ (inégalité de —), 106.
Semi-logarithmique (échelles —s), 72.
Somme (d'aléas numériques), 243.
Support (d'une mesure), 194.

Triangle (inégalité du —), 107, 119.
Type (d'ordre), 26.

Uniforme (mesure —), 189.

Variable (aléatoire), 192, 199.
Variance, 100, 128, 270.

TABLE DES MATIÈRES

PAGES

AVANT-PROPOS 1

SOMMAIRE 3

INSTRUCTIONS DE LECTURE POUR LES ÉTUDIANTS..... 5

BIBLIOGRAPHIE 7

PREMIÈRE PARTIE

LES NOMBRES

CHAPITRE PREMIER. — *Nombres, ordres, approximations*..... 11

1. Ensembles de nombres 11
2. Entiers naturels : **N** 14
3. Entiers positifs et négatifs : **Z** 16
4. Les rationnels : **Q*** et **Q** 19
5. Les réels : **R** et **R*** 22
6. Les décimaux : **D** 23
*7. Types d'ordre (échelles) : la comparaison..... 26
8. Approximations (par J.-L. PIEDNOIR)..... 29
9. De quels nombres faut-il se servir ?..... 34

Récapitulation 37

CHAPITRE II. — *Fonctions numériques* (avec la collaboration de J.-L. PIEDNOIR) 38

1. La fonction linéaire 38
1 *bis*. Fonctions affines (progressions arithmétiques) 44

PAGES

2. La fonction exponentielle 46
3. La fonction logarithme 56
4. La fonction puissance 62
*5. Les quatre types fondamentaux de fonctions numériques 66
6. Application : les papiers fonctionnels 69
7. Calcul pratique de logarithmes 73
Récapitulation 77

DEUXIÈME PARTIE

LA MESURE EN DESCRIPTION STATISTIQUE

CHAPITRE III. — *Produit scalaire* 81

1. Produit scalaire 82
*2. Caractérisation algébrique du produit scalaire 88
3. Longueur d'un vecteur dans $\mathbf{R}^n$ 91
4. Vecteurs orthogonaux 94
5. Application à la description statistique : covariance, variance 97
6. Inégalité de Schwarz. Inégalité du triangle 105
Récapitulation 109

CHAPITRE IV. — *Mesures de la distance et de la dispersion* ... 110

1. Longueur et distance euclidiennes 110
2. Moyenne et distance minimum 121
3. Mesure de la dispersion : variance et écart type 126
*4. Métriques non linéaires 130
Récapitulation 136

CHAPITRE V. — *Cosinus et corrélation linéaire* 138

1. Cosinus de deux vecteurs 138
2. Cosinus et ajustement linéaire 145
3. Coefficient de corrélation linéaire et fonctions affines .. 152
Récapitulation 159

TROISIÈME PARTIE

LA MESURE EN PROBABILITÉS

PAGES

CHAPITRE VI. — *Mesures de probabilité* 163
1. Possibilités et probabilités 163
2. Mesures de probabilité 170
*3. Identité des mesures, différences et axiomes des probabilités 178
4. Distributions de probabilités et construction des mesures de probabilité 182
Récapitulation 190

CHAPITRE VII. — *Aléas* 192
1. Aléas 192
2. Aléa défini par une application 195
3. Aléa de Bernoulli comme aléa image 200
4. Aléas hypergéométriques (tirages sans remise) 201
5. Aléas de pile ou face 206
6. Aléa binomial (tirages avec remise) 209
7. Aléas quotients 209
Récapitulation 211

CHAPITRE VIII. — *Probabilités conditionnelles* 213
1. Mesures de probabilités conditionnelles 213
2. Produit cartésien et croisement d'aléas 222
3. Aléas et événements indépendants en probabilité 231
Récapitulation 238

CHAPITRE IX. — *Aléas numériques* 241
1. Les aléas numériques et l'algèbre linéaire 241
2. Espérances 249
3. Covariance, variance 264
4. Première loi des grands nombres. Inférence statistique 275

INDEX DES SYMBOLES ET NOTATIONS 283
INDEX DES TERMES 285

SUP

SECTION « LE PSYCHOLOGUE »

1	ANDRÉ REY	**L'EXAMEN CLINIQUE EN PSYCHOLOGIE**
2	JEAN-MARIE FAVERGE, JACQUES LEPLAT et BERNARD GUIGUET	**L'ADAPTATION DE LA MACHINE A L'HOMME**
3	CHARLES NAHOUM	**L'ENTRETIEN PSYCHOLOGIQUE**
4	MAURICE COUMÉTOU	**LES EXAMENS SENSORIELS**
5	GASTON VIAUD	**LES INSTINCTS**
6	GABRIEL DESHAIES	**PSYCHOPATHOLOGIE GÉNÉRALE**
7	CHARLES CHANDESSAIS	**LA PSYCHOLOGIE DANS L'ARMÉE**
8	SUZANNE PACAUD	**LA SÉLECTION PROFESSIONNELLE**
9	DIDIER ANZIEU	**LES MÉTHODES PROJECTIVES**
10	GEORGES BASTIN	**LES TECHNIQUES SOCIOMÉTRIQUES**
11	RÉMY CHAUVIN	**LE COMPORTEMENT SOCIAL CHEZ LES ANIMAUX**
12	JEAN-FRANÇOIS LE NY	**LE CONDITIONNEMENT**

STÉPHANE EHRLICH et CLAUDE FLAMENT

13 **PRÉCIS DE STATISTIQUE**

PAUL CHAUCHARD

14 **DES ANIMAUX A L'HOMME**
(Psychismes et cerveaux)

HENRI PIÉRON

15 **EXAMENS ET DOCIMOLOGIE**

HENRI EY

16 **LA CONSCIENCE** (2e éd.)

PIERRE OLÉRON

17 **LES ACTIVITÉS INTELLECTUELLES**

IRÈNE LÉZINE

18 **PSYCHOPÉDAGOGIE DU PREMIER AGE**

DONALD E. SUPER

19 **LA PSYCHOLOGIE DES INTÉRÊTS**

ROGER PIRET

20 **PSYCHOLOGIE DIFFÉRENTIELLE DES SEXES**

JOSEPH NUTTIN

21 **LA STRUCTURE DE LA PERSONNALITÉ**

FRANÇOIS GAUCHET, ROGER LAMBERT et JACQUES VIOLET

22 **LE CALCUL AUTOMATIQUE EN PSYCHOLOGIE**

JEAN DELAY

23 **LES MALADIES DE LA MÉMOIRE**

PIERRE FOUILHÉ

24 **LA PSYCHOLOGIE COMMERCIALE**

POL DEBATY

25 **LA MESURE DES ATTITUDES**

MAURICE DE MONTMOLLIN

26 **LES SYSTÈMES HOMMES-MACHINES**
(Introduction à l'ergonomie)

JEAN-MARIE FAVERGE

27 **PSYCHOSOCIOLOGIE**
DES ACCIDENTS DU TRAVAIL

ROBERT FRANCÈS

28 **PSYCHOLOGIE DE L'ESTHÉTIQUE**

STANISLAW TOMKIEWICZ

29 **LE DÉVELOPPEMENT BIOLOGIQUE**
DE L'ENFANT

MARC BARBUT

30 **MATHÉMATIQUES**
DES SCIENCES HUMAINES
I. — Combinatoire et algèbre

ANDRÉ BERGE

31 **LES PSYCHOTHÉRAPIES**

DIDIER ANZIEU et JACQUES-YVES MARTIN

32 **LA DYNAMIQUE**
DES GROUPES RESTREINTS

MARC BARBUT

33 **MATHÉMATIQUES**
DES SCIENCES HUMAINES
II. — Nombres et mesures

PAUL GUILLAUME

34 **LA FORMATION DES HABITUDES**

B. ANDREY et J. LE MEN

35 **LA PSYCHOLOGIE A L'ÉCOLE**

VOLUMES EN PRÉPARATION

PAUL ALBOU
LES QUESTIONNAIRES PSYCHOLOGIQUES

PIERRE GOGUELIN
LA FORMATION PSYCHO-SOCIALE DANS L'INDUSTRIE

PAUL GUILLAUME
L'IMITATION CHEZ L'ENFANT

CARL HOYOS
PSYCHOLOGIE DE LA CIRCULATION ROUTIÈRE

BÄRBEL INHELDER
LES STADES EN PSYCHOLOGIE DE L'ENFANT

THÉRÈSE LEMPÉRIÈRE
LES GRANDES MALADIES MENTALES

THÉRÈSE LEMPÉRIÈRE
SÉMÉIOLOGIE DES TROUBLES MENTAUX

SUZANNE PACAUD
PSYCHOLOGIE DU VIEILLISSEMENT

MAURICE REUCHLIN
LA PSYCHOLOGIE DIFFÉRENTIELLE

JEAN STOETZEL et ALAIN GIRARD
LES SONDAGES D'OPINION PUBLIQUE

DAVID VICTOROFF
PSYCHOLOGIE DE LA PUBLICITÉ

1968-1. — Imprimerie des Presses Universitaires de France. — Vendôme (France)
ÉDIT. N° 29 859 IMPRIMÉ EN FRANCE IMP. N° 20 621

LES PRÉCIS DE L'ENSEIGNEMENT SUPÉRIEUR

LE PSYCHOLOGUE
Section dirigée par Paul FRAISSE

L'ÉDUCATEUR
Section dirigée par Gaston MIALARET

LE SOCIOLOGUE
Section dirigée par Georges BALANDIER

LE LINGUISTE
Section dirigée par André MARTINET

L'HISTORIEN
Section dirigée par Roland MOUSNIER

LE GÉOGRAPHE
Section dirigée par Pierre GEORGE

L'ÉCONOMISTE
Section dirigée par Pierre TABATONI

LE MATHÉMATICIEN
Section dirigée par Jean-Pierre KAHANE

LE PHYSICIEN
Section dirigée par Hubert CURIEN

LE CHIMISTE
Section dirigée par Jacques BÉNARD

PAIDEIA
Bibliothèque pratique
de Psychologie et de Psychopathologie de l'Enfant
dirigée par Georges HEUYER

LES PRÉCIS DES CLASSES SUPÉRIEURES

INITIATION PHILOSOPHIQUE
Section dirigée par Jean LACROIX

LES GRANDS TEXTES
Bibliothèque classique de Philosophie
dirigée par Claude KHODOSS et Jean LAUBIER

PHILOSOPHES

29 859. C

LIE MINERVE
NEUF • OCCASION
79, BOUL ST MICHEL
PARIS Ve